REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-04-

0347

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE May 31, 2004	3. REPORT TYPE AND DATES COVERED Final December 2000 - February 2004

4. TITLE AND SUBTITLE Intelligent Information Systems Institute	5. FUNDING NUMBERS F49620-01-1-0076
6. AUTHOR(S) Gomes, Carla	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cornell University 5133 Upson Hall Ithaca, NY 14853	8. PERFORMING ORGANIZATION REPORT NUMBER 39383
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 4015 Wilson Boulevard, Room 824A Arlington, VA 22203-1954	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; distribution is Unlimited	12b. DISTRIBUTION CODE U

13. ABSTRACT *(Maximum 200 Words)*

This report describes the research accomplishments of the Intelligent Information Systems Institute (IISI) at Cornell during the first three years of operation. IISI's mandate is threefold: To perform and stimulate research in computational and data-intensive methods for intelligent decision making systems; to foster collaborations between Cornell researchers, AFRL, in particular IF, and the scientific community; and to play a leadership role in the research and dissemination of the core areas of the institute. IISI's research spans across various areas and disciplines, organized into four themes: Theme 1 - Information Capture and Discovery from heterogeneous information sources; Theme 2 - Controlling Computational Cost; Theme 3 – Pervasive computing: autonomous distributed agents and communication networks; and Theme 4 – Advanced Computing Architectures. In its first three years of existence IISI has successfully fostered collaborations between Cornell's researchers, AFRL/IF researchers, and the research community at large in the area of computing and information science. AFRL researchers have an active interaction with Cornell researchers and participate in Cornell research projects, facilitating technology transfer from academia into the military arena. Cornell researchers on the other hand have had the opportunity to be exposed to interesting Air Force related problems.

14. SUBJECT TERMS	15. NUMBER OF PAGES 60
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT U	18. SECURITY CLASSIFICATION OF THIS PAGE U	19. SECURITY CLASSIFICATION OF ABSTRACT U	20. LIMITATION OF ABSTRACT UU

20040709 036

Intelligent Information Systems Institute
AFOSR Grant F49620-01-1-0076
Final Report
May 2004

Carla P. Gomes,
Computing and Information Science
Cornell University
Ithaca, NY

(607) 255-9189 (phone)
(607) 255-4428 (fax)
gomes@cs.cornell.edu

1. Executive Summary

This report describes the research accomplishments of the Intelligent Information Systems Institute (IISI) at Cornell during the first three years of operation. IISI's mandate is threefold: To perform and stimulate research in compute- and data-intensive methods for intelligent decision making systems; to foster collaborations between Cornell researchers, AFRL, in particular IF, and the scientific community; and to play a leadership role in the research and dissemination of the core areas of the institute.

IISI was modeled after national research institutes such as the DIMACS center for Discrete Mathematics. The institute promotes research collaborations with its sponsors and the research community at large. The activities supported by the institute include research collaborations and projects, visiting scientists, working groups, conferences and workshops, special programs on specific topics and challenge problems, technical reports, and other publications. The IISI director and principal investigator is Carla P. Gomes. IISI's advisory board includes Robert Constable, Dean for the Office of Computing and Information Science at Cornell University, Northrup Fowler, III, acting Chief Scientist, AFRL Information Directorate, Charles Messenger, Branch Chief, AFRL Information Directorate Information Technology division, Neal Glassman AFRL/ AFOSR-EOARD, and Maj. Juan Vasquez AFRL/AFOSR.

The IISI supports basic research within the Faculty of Computing and Information Science (FCIS) and promotes a cross fertilization of approaches from different disciplines, including Operations Research, Mathematics, Statistics and Physics. IISI's research spans across various disciplines and topics, organized into four themes: Theme 1 - Information Capture and Discovery from heterogeneous information sources; Theme 2 - Controlling Computational Cost; Theme 3 – Pervasive computing: autonomous distributed agents and communication networks; and Theme 4 – Advanced Computing Architectures. These central themes cover several research areas, namely: Search and Complexity; Combinatorial Optimization; Planning and Scheduling; Knowledge Representation; Data Mining and Information Retrieval; Reasoning under Uncertainty; Natural Language Processing; Machine Learning, and Agent Technology.

2. Research Accomplishments – Cornell Researchers

In its first three years of existence IISI has successfully fostered collaborations between Cornell's researchers, AFRL/IF researchers, and the research community at large in the area of computing and information science. AFRL researchers have an active interaction with Cornell researchers and participate in Cornell research projects, facilitating technology transfer from academia into the military arena. Cornell researchers on the other hand have had the opportunity to be exposed to interesting Air Force related problems.

The research focus of the Intelligent Information Systems Institute (IISI) is on the development and analysis of computational and data intensive methods for intelligent distributed decision making systems. These problems are ubiquitous and occur in areas as diverse as planning, scheduling, learning, diagnosis, verification, reasoning, and protein folding. The core research areas of the institute include knowledge representation and reasoning, search and complexity, uncertainty, distributed problem solving and agent technology, databases and data mining, machine learning, and information retrieval and natural language processing. To better understand and tame such challenge problems, we promote a cross fertilization of approaches from different disciplines such as computer science, engineering, operations research, mathematics, and physics, which is reflected in the diversity of the departmental affiliation of our members.

Below we highlight our research accomplishments and findings, considering four major research themes:

> Theme 1 - Mathematical and Computational Foundations of Complex Networks
> Theme 2 - Controlling Computational Cost
> Theme 3- Pervasive computing: autonomous distributed agents and communication networks
> Theme 4 – Advanced Computing Architectures

Theme 1 - Mathematical and Computational Foundations of Complex Networks

John Hopcroft is leading a research project, together with Bart Selman, on foundational work on information capture and knowledge discovery using <u>ultra large</u> data repositories, from heterogeneous sources. Cornell students Omar Kahn, Brian Kulis, and Justin Yang and researcher Nate Gemelli, from AFRL/IF, are members of Hopcroft's research team.

Emergent properties of large linked networks have recently become the focus of intense study. This research is driven by the increasing complexity and importance of large networks, such as the World Wide Web, the electricity grid, and large social networks that capture relationships between individuals. Real-world networks generally exhibit properties that lie somewhere in-between those of highly structured networks and purely random ones. So far, most research has focused on using static properties, such as the connectivity of the nodes in the network and the average distance between two nodes, to explain the complex structure. However, these networks generally evolve over time and so temporal characteristics are a key source of interest.

The goal of Hopcroft's team is to provide techniques for the study of the dynamics and evolution of large linked networks. Hopcroft et al. are also investigating the problem of discovering hidden structure from large databases. They are interested in whether one can determine, e.g., by examining the link structure of the bibliographic citations over time, when a new community forms. In particular they are interested in early detection of a new community before it is obvious. More specifically, they are interested in tracking changes in large-scale data by periodically creating an agglomerative clustering and examining the evolution of clusters (communities) over time. For their empirical work they examine a large real-world data set: the NEC CiteSeer database, a linked network of over 250,000 papers. Tracking changes over time requires a clustering algorithm that produces clusters stable under small perturbations of the input data. However, small perturbations of the CiteSeer data lead to significant changes to most of the clusters. One reason for this is that the order in which papers within communities are combined is somewhat arbitrary. However, certain subsets of papers, called natural communities, correspond to real structure in the CiteSeer database and thus appear in any clustering. By identifying the subset of clusters that remain stable under multiple clustering runs, they get the set of natural communities that can be tracked over time. They have demonstrated that such natural communities allow one to identify emerging communities and track temporal changes in the underlying structure of our network data.

Hopcroft et al.'s work provides a framework for studying the temporal evolution of the community structure of large linked networks. They have introduced the notion of natural communities that can be used to identify a relatively stable core of a hierarchical agglomerative clustering. Their approach exploits the inherent instabilities in clusterings in high-dimensional spaces. The true structure in the data is revealed by averaging out the large number of ``accidental" clusters that emerge in any single clustering run. In experiments on the CiteSeer network, they have shown how the natural communities can be used to study the evolution of the network by tracking established communities and uncovering new, emerging community structure. The next step is to evaluate the approach on other evolving linked networks.

In the context of developing an information extraction (IE) system in the domain of natural disasters, Cardie and Vincent (Cardie 2002 and Vincent and Cardie 2002) developed an approach to noun phrase coreference resolution that outperforms all existing approaches to date. By way of background, noun phrase coreference resolution refers to the problem of determining which noun phrases (NPs) refer to each real-world entity mentioned in a document; it has been identified as arguably the most critical outstanding problem to developing information extraction systems that can operate at useable levels of accuracy and recall.

Cardie and Vincent have made substantial progress towards the objective listed above. In particular, they have produced a noun phrase coreference system that relies on standard machine learning methods and that outperforms all existing approaches to the problem to date as evaluated on the two widely available coreference data sets. There is still substantial room for improvement, however, and so work will continue on this problem.

The ultimate goal of natural language processing is to enable computers to use human language as a communication medium both robustly and gracefully. However, because of the subtleties of human language, this goal cannot be achieved without access to large quantities of high-quality linguistic and domain knowledge. A major focus of Lee's research is the development of "knowledge-lean" methods to overcome this knowledge acquisition bottleneck. She is investigating techniques that allow a system to automatically learn linguistic and domain knowledge directly from text. A major focus of Lee's work has been the study of distributional similarity and distributional clustering. Lee's work explores the use of distributional similarity as a powerful tool for bootstrapping the knowledge acquisition process. The underlying idea is quite intuitive: information about an object can be gleaned from the objects that are similar to it, where similarity can be computed from unannotated data alone. But there is a multitude of ways to implement this idea; she is currently investigating different similarity-based paradigms, both theoretically and through experiments on large datasets.

Also, work with Rie Kubota Ando has analyzed Latent Semantic Indexing, a commonly-used and highly influential information retrieval technique that seeks to uncover hidden document relationships without engaging in standard natural language processing. They have shown that the performance of this algorithm can be tied to the uniformity of the underlying topic-document distribution, and further has provided a new algorithm, which automatically compensates for distributional non-uniformity.

Lee and her colleagues have also considered applications ranging from finding word boundaries in streams of Japanese to creating English versions of computer-generated mathematical proofs.

Thorsten Joachims joined the IISI in 2002 after joining the Department of Computer Science as an Assistant Professor late in 2001. His research interests center on a synthesis of theory and system building in the field of machine learning, with a focus on support vector machines, text-mining, and machine learning in information access.

Most machine learning research over the last decade has focused on problems like classification and regression, where the prediction is a single univariate variable. But what if someone needs to predict complex objects like trees, sequences, or orderings? Such problems arise, for example, when a natural language parser needs to predict the correct parse tree for a given sentence, when a navigation assistant needs to predict the route a user prefers for getting to the destination, or when a search engine needs to predict which ranking is best for a given query. They explored new methods for predicting such complex objects. In particular, they developed support vector approaches that cover some of these problems. They generalize the idea of margins to complex prediction problems and a large range of loss functions. While the resulting training problems have exponential size, they proposed a simple algorithm that allows training in polynomial time. Empirical results show that these new methods outperform existing techniques on a variety of problems.

Another area of machine learning where they contributed results is transductive learning. Consider the following type of prediction problem: you have a large database of object (e.g. text documents) that you would like to organize according to some classification. You know the classification for a small subset of the objects and you would like to infer the classes for the remaining objects. A typical example is relevance feedback in information retrieval: a user specifies some documents as relevant/non-relevant and wants to find all other relevant documents in the collection. This type of machine learning task is called transductive inference. A key difference to the typical inductive machine learning problem is that the location of the test points can be observed by the learning algorithm. They showed that exploiting cluster structure in the test set can help make more accurate predictions, in particular for text classification. However, designing transductive learning algorithms that simultaneously minimize training error and maximize a measure of clustering in the test set is computationally challenging. To overcome this problem, they presented an approach based on spectral graph partitioning. In this approach, objects in the database form the nodes of a graph, while the edges represent dependencies. For such graphs, they generalized ratio-cuts to find a clustering of the database that obeys the known classifications for the training examples. The relaxation of the resulting optimization problem can be solved as an eigenvalue problem. Empirical results show improved prediction accuracy especially for small training sets. Furthermore, the framework is very flexible, making it possible to implement co-training as a special case.

Their contributions in the area of information retrieval explore the use of implicit feedback for improving search engines. A central goal of information retrieval is the design of functions that rank documents according to their relevance to a query. They developed an approach to automatically learning such ranking functions by phrasing this task as learning a parameterized ordering over a finite domain. Taking an empirical-risk-minimization approach with Kendall's Tau as the loss function, they developed a Support Vector algorithm that leads to a convex training problem. An important property of this algorithm is that it can be trained with partial information like "for query Q, document A should be ranked higher than document B". Such feedback can be inferred more easily from implicit feedback (e.g. the clicking behavior of users) than traditional absolute feedback of the form "for query Q, document A is relevant/non-relevant". Experiments show that the method can use clickthrough data to effectively adapt the retrieval function of a meta-search engine to a particular group of users, outperforming Google in terms of retrieval quality after only a couple of hundred training queries.

Data mining is one of the very promising information technologies today. The Cornell Database Group has developed some of the fastest known algorithms for several data mining problems such as classification and regression tree construction, finding maximal large itemsets, and mining sequential patterns. The objective of this effort is to leverage previous experience to more sophisticated data mining problems as they might occur in real-life applications encountered by the Air Force, such as intrusion detection or interactive manipulation of large data sets.

Caruana's research is in machine learning and data mining. His current focus is on inductive transfer, learning rankings, adaptive clustering, and applications of these methods to problems in medical decision-making and bioinformatics. Inductive transfer is a subfield of machine learning that achieves better performance by learning to solve many related problems simultaneously: often it is easier to learn 100 related problems at the same time than to learn any one of them in isolation because what is learned for one problem often is useful for another problem and can be transferred. Learning to order things is an exciting new area in machine learning that has important applications in information retrieval, bioinformatics, and medicine. Caruana is developing algorithms that learn rankings for problems in medical decision making where it may be difficult to assess absolute risk for a patient, but easier to learn to order patients by relative risk. He developed the first machine learning algorithm for learning rankings with neural nets (RankProp) in 1996. Caruana's work in clustering is a recent focus for him. His interest in clustering arose from problems he discovered when applying traditional clustering methods to protein folding with colleagues in bioinformatics. A theme that runs through all of Caruana's work is the importance of developing methods that are applicable to and effective on real-world problems. He likes to mix algorithm development with applications work to insure that the methods he develops are useful. Below is more detailed information on three very successful ongoing projects led by Caruana: Ensemble Selection, Meta Clustering, and Probability Calibration.

ENSEMBLE SELECTION:

An ensemble is a collection of models whose predictions are combined by weighted averaging or voting. A necessary and sufficient condition for an ensemble to be more accurate than any of its individual members is if the models are accurate and diverse (Dietterich 1999). Many methods have been proposed to generate accurate, yet diverse, sets of models: bagging, boosting, error correcting codes, feature boosting, ... Instead of generating many models using a single learning algorithm, ensemble selection generates a diverse set of models using many different algorithms: Support Vector Machines (SVMs), artificial neural nets, k-nearest neighbor, decision trees, bagged decision trees, boosted decision trees, and boosted stumps. For each algorithm we train models using many different parameter settings. For example, we train 121 SVMs by varying the margin parameter C, the kernel, and the kernel parameters (e.g. varying gamma with RBF kernels.)

In total, they train about 2000 models for each problem. Rather than combine all of these models, some good and some bad in an ensemble, they use forward stepwise selection from the library of models to find a subset of models that when averaged together yield excellent performance. The basic selection procedure is very simple:

- Start with the empty ensemble.

- Add to the ensemble the model in the library that maximizes the ensemble's performance to the error metric on a hillclimb (validation) set.
- Repeat Step 2 for a fixed number of iterations or until all the models have been used.
- Return the ensemble from the nested set of ensembles that has maximum performance on the hillclimb (validation) set.

Models are added to an ensemble by averaging their predictions with the models already in the ensemble. This makes adding a model to the ensemble very fast, allowing ensembles with excellent performance to be found in minutes from libraries with 2000 models. Most importantly, this selection procedure allows the ensemble to be optimized to any easily computed performance metric. This is the first supervised learning method they know of that has this flexibility. So far they have evaluated ensemble selection on ten different performance metrics, making this one of the most comprehensive empirical evaluations of a learning algorithm. Because ensemble selection generates so many different models, libraries usually contain a few models with excellent performance on any given performance metric. Just selecting the best single model from the library yields state-of-the-art performance. In experiments on seven different test problems, ensemble selection consistently finds ensembles that outperform the best models that can be trained with any competing learning method. This suggests that using different learning methods and parameter settings for these methods is an effective procedure for generating a diverse set of good-performing models.

META CLUSTERING:

In supervised learning, any sufficiently accurate model is suitable for most uses. This is not true in clustering: a clustering that is perfect for one use can be inadequate for another use. Yet most clustering algorithms search for one optimal clustering of the data. This optimal clustering often is not well defined. The ``best" clustering depends on what it will be used for. Data may need to be clustered in different ways for different uses. For example, a clustering appropriate for consumer marketing is not likely to be appropriate for medical research, and vice-versa. An analogy will help explain how meta clustering will help users find the best clustering for their purpose. Photoshop, the photo editing software, has a tool called ``variations" that presents to users different transformations of an image with various color balances, brightnesses, and contrasts. Instead of having to know the correct tool to use to improve their picture, the user just selects the variation that looks best. The selected variation becomes the new center, and variations of it are presented, allowing users to quickly zero in on the desired rendition. The goal of meta clustering is to provide a similar ``variations" tool for clustering so that users do not have to know how to modify clustering distance metrics and clustering algorithms to achieve useful clusterings of the data. Instead, users will be presented with an organized set of clustering variations, and will be able to select and refine the clustering variation(s) that are best suited to their purposes. The IISI-sponsored research on meta clustering has focused on developing a stochastic algorithm for generating many alternate, yet high quality, clusterings, and on how to cluster these clusterings at the meta level so that the user sees an organized collection of alternate clusterings. The methods they have developed have been developed while working on a text clustering problem.

PROBABILITY CALIBRATION:

A model is well calibrated if when it predicts that a set of cases have a probability of p to be class 1, about p percent of those models actually turn out to be class 1, for all probabilities p between 0 and 1. As part of the research above in ensemble selection they discovered something surprising about model calibration: although single decision trees typically do not predict well calibrated probabilities, bagged decision trees typically yield well calibrated probabilities, yet boosting the same decision trees yields very poorly calibrated predictions. This is surprising because bagging and boosting both are ensemble methods that average the predictions from multiple trees. The experiments also showed that boosted trees yield excellent performance on other performance metrics such as accuracy and ROC Area that do not need well calibrated predictions. SVMs also have good performance on these non-calibration metrics, but must have their predictions (which are unbounded) scaled to the interval 0-1 if they will be interpreted as probabilities. A method called Platt Scaling is one way to scale SVM predictions to good probabilities. They wondered if Platt scaling might transform the poor probabilities from boosted decision trees (which already are on 0-1 and thus don't necessarily need any scaling) to predictions that were better calibrated. Preliminary experiments suggest that the answer is a definite yes: Platt scaling is able to transform boosting's poor probabilistic predictions into excellent probabilities, thereby transforming boosted decision trees into a true general purpose learning method that has excellent performance on all metrics on which they have tested it.

Gehrke's group focuses on constrained-based mining for questions such as "find all frequent itemsets where the total duration is at least 50 minutes. Finding itemsets with constraints has wide applications, such as in intrusions detection, web log mining, fraud detection, classification, association rules, and collaborative filtering. The problem can be stated abstractly as follows. Let M be a finite set of items from some domain (for example, products in a grocery store). All the items have a common set of descriptive attributes (i.e., the name, brand, or price of the item). A predicate (or constraint) over a set of items is a condition that the set has to satisfy. The goal of constraint-based market basket analysis is then: Given a set of items M, a set of predicates P1, P2, ..., Pn, find all subsets of M that satisfy P1 and P2 and ... and Pn.

Important classes of constraints, most notably monotone and antimonotone, have been studied as important classes of constraints. There exist algorithms that take advantage of each class of constraints. However, their main deficiency is that they each handle only one class of constraints efficiently. More recently, Raedt and Kramer have generalized these algorithms to allow several types of constraints, but this generalization only handles one type of constraint at a time. With funding from the IISI, they have developed a novel algorithm called DualMiner, which efficiently mines constraint-based itemsets by simultaneously taking advantage of both monotone and antimonotone predicates. They developed the algorithm and complement a theoretical analysis and proof of correctness of DualMiner with an experimental study that shows the efficacy of DualMiner compared to previous work; DualMiner outperforms previous work up to several orders of magnitude.

Even with DualMiner, the field of mining itemsets with constraints was still an unconnected collection of algorithms, and for each new type of constraint, a new algorithm needed to be developed. For example, there exist algorithms that mine monotone or anti-monotone constraints, or mine convertible constraints together with either monotone or anti-monotone constraints, but there does not yet exist a unified algorithm that can mine all of three types of constraints together.

In follow-up work over the last year, they developed a unified framework for constrained itemset mining that applies to any type of constraint. Their framework is based on the concept of efficiently finding a witness, which is a single itemset X on which we can test whether the constraint holds. This test will provide information about properties of other itemsets. This information can then be used for pruning the search space. The notion of a witness has conceptual implications. For example, they now can efficiently mine all three types of constraints *simultaneously* (by finding witnesses for each constraint), and we can also mine complicated constraints that are neither monotone, antimonotone, nor convertible. As a demonstration, they will introduce an efficient algorithm for finding a witness for constraints involving the variance of a set of items, a problem that has been open for several years in the data mining community.

Ben-David's research in the past academic year focused on three issues: analysis of the possible success guarantees that may be proved for Kernel based learning, developing a Machine Learning approach to the task of Data Integration over disparate data sources, and pushing forward the theory of Learning To Learn.

1) "A Priori Generalization Bounds for Kernel Based Learning" (invited to a special issue of the Journal of Machine Learning Research).

Kernel based learning methods are among the most powerful and widely used learning paradigms. A large body of research supports the algorithmic and computational complexity aspects of kernel based learning, and in particular of the support vector machines approach. When it comes to the information theoretic aspects, namely figuring out needed sample sizes and generalization guarantees, our understanding of these methods seems to be lacking. Most (if not all) of the existing generalization results have an a posteriori nature - the bounds are based on viewing the actual input training data, rather than on some a prior assumptions about the nature of the learning problem. As a contrasting example, consider the PAC learning framework, or the empirical risk minimization approach to agnostic learning, where the fundamental generalization results are of the form "if the target function (or labeling distribution) is close enough to a member of our hypothesis class, then with high probability, if we observe so many labeled examples, our hypothesis will have error smaller than ...". It would be desirable to develop an 'a priori based' generalization theory for kernel based methods as well.

A first step in this direction was taken in a recent paper by Ben-David, Eiron and Simon. It is shown there that most of the finite concept classes of any fixed VC-dimension cannot be embedded into classes of Euclidean half-spaces unless the resulting margins are very small and the Euclidean dimension of the half-spaces is very large. These results imply that the known generalization bounds that are based on either VC-dimension or margins cannot provide a priori generalization guarantees for kernel based learning (even for classes on which, without embedding them into half-spaces, empirical risk minimization does enjoy such guarantees).

While these results rule out the use of the most commonly used theoretical generalization bounds (those based on either margins or VC - dimension), they do not rule out the possibility that generalization may be guaranteed via different considerations. The most natural candidate to investigate next is the notion of *sparsity*. It can be shown that, given a labeled training sample, the generalization of a hypothesis that is consistent with this sample may be guaranteed in terms of the *'sparsity'* (or the representation size within some fixed representation scheme) of this hypothesis.

In this work Ben-David analyze the suitability of sparsity for the purpose of providing a priori generalization bounds for learning via kernel based methods.

2) "A Theoretical Framework for Learning from Disparate Data Sources" (Joint work with Johannes Gehrke and Reba Schuller, published in the proceeding of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining).

Many enterprises incorporate information gathered from a variety of data sources into an integrated input for some learning task. For example, aiming toward the design of an automated diagnostic tool for some disease, one may wish to integrate data gathered in many different hospitals. A major obstacle to such endeavors is that different data sources may vary considerably in the way they choose to represent related data. In practice, the problem is usually solved by a manual construction of semantic mappings and translations between the different sources. Recently there have been attempts to introduce automated algorithms based on machine learning tools for the construction of such translations.

In this work Ben-David et al. propose a theoretical framework for making classification predictions from a collection of different data sources, without *creating explicit translations between them.* The framework allows a precise mathematical analysis of the complexity of such tasks, and it provides a tool for the development and comparison of different learning algorithms. The main objective, at this stage, is to demonstrate the usefulness of computational learning theory to this practically important area and to stimulate further theoretical and experimental research of questions related to this framework.

3) "Exploiting Task Relatedness for Multiple Task Learning" (Joint work with Reba Schuller, submitted to the upcoming NIPS'02 conference)

The approach of learning of multiple "related" tasks simultaneously has proved quite successful in practice; however, theoretical justification for this success has remained elusive. The starting point of previous work on multiple task learning has been that the tasks to be learned jointly are somehow "algorithmically related", in the sense that the *results* of applying a specific learning algorithm to these tasks are assumed to be similar. Ben-David et al. take a logical step backward and offer a data generating mechanism through which our notion of task-relatedness is defined. They provide a formal framework for task relatedness that captures a certain sub-domain of the wide scope of issues in which one may apply a multiple task learning approach. Their notion of similarity between tasks is relevant to a variety of real life multi-task learning scenarios and allows the formal derivation of strong generalization bounds (bounds that are strictly stronger than the previously known bounds for both the learning-to-learn and the multi-task-learning scenarios). They provide general conditions under which our bounds guarantee smaller sample size per task than the known bounds for the single task learning approach.

Theme 2 - Controlling Computational Cost

This theme encompasses foundational work on “typical” computational complexity, study of structure, and randomization in hard computational problems. This research combines formal analysis and design of optimization techniques with the study of a range of applications to large-scale computational problems, such as distributed networks, autonomous agents, and combinatorial auctions, applications such as planning and scheduling, autonomous distributed agents, and combinatorial auctions. Central themes of our work are (1) the integration of optimization concepts from artificial intelligence, and operations research, in particular mathematical programming and constraint programming, (2) foundational work on “typical” computational complexity, (3) the study of the impact of structure on problem hardness of hard computational problems, (4) the use of randomization techniques to improve the performance of search methods, and (5) global optimization through machine learning and response surface methods. Below some findings are highlighted.

In complex decision problems arising in military applications, the decision maker is expected to make sensitive decisions in the face of large volumes of continuously arriving information, multi-parametric and dynamic environment, time limitations, and both adversarial and cooperating agents. Attempting to replace such decision makers with automated agents would be highly presumptuous and unrealistic. Yet, it is believed that we can provide these decision makers with tools that would support their decision making process. For instance, we can help them organize incoming information by filtering it and focusing their attention on the more relevant and urgent data and presenting it appropriately. We can analyze information and alert them if we recognize problems or opportunities. We can propose good courses of actions, and help evaluate plans the decision maker may be contemplating. Any such tool that enhances the productivity or the quality of decisions made across different ranks is likely to have a very significant impact on the overall performance of the military forces.

The main objective of the long term research of Carmel Domshlak has been providing theoretical and practical foundations for such decision support systems. He has focused on exploiting and extending the recent achievements of qualitative decision theory, examining how the new formalisms emerging from qualitative decision theory research can be used in various decision support applications, and extending theoretical understanding of these formalisms, their expressive power, algorithmic complexity, and relation to canonical (yet hardly practical) quantitative models of objectives originated in the field of mathematical economics. In particular, Carmel Domshlak and his colleagues have been developing a novel methodology combining the advantages of both qualitative and quantitative models [BDK04,DL04]. The idea behind this research is *to use a unifying quantitative model to reason and represent heterogeneous qualitative preference statements*, by this combining the relative advantages of qualitative and quantitative methods.

Qualitative statements capture the type of preference information human decision makers feel comfortable providing. These statements include general statements about preferred values/results, about the relative importance of different attributes of a solution/plan, and explicit ranking of a small set of proposed alternatives. However, recent results by Carmel Domshlak and his colleagues showed that directly reasoning with even homogeneous sets of such statements can be computationally prohibitive [BBDHP04a,Lang02]. Thus, developing specialized mechanisms for reasoning about heterogeneous sets of qualitative statements is likely to be even more difficult. On the other hand, quantitative methods, and utility theory in particular, are convenient to reason with. Mathematically, they are quite simple and they embody rich information about preferences. Yet, direct elicitation of utility functions does not appear to be an option for practical decision support tools. Fortunately, it turns out that many preference statements can be interpreted as constraints on utility functions. Thus, utility functions, in a sense, provide a unifying semantics for qualitative preference statements, allowing us to combine heterogeneous statements in a single, computationally attractive formalism.

The core element of the methodology developed by Carmel Domshlak is the idea of *knowledge compilation*, effectively exploiting various tools of both linear and non-linear mathematical programming. Informally, the purpose of the knowledge compilation module is to generate a good numerical approximation of a given set of qualitative decision-making guidelines. Carmel Domshlak developed a new representation theory for generalized additive value functions [BDK04]. Specifically, he provided conditions under which there exists a particular factored value function representing the user's preference relation. These representation theorems show that certain partial orders induced by sets of qualitative statements of conditional preference for variable values and conditional relative importance between different variables, can be represented using a compact generalized additive value function. This result is much stronger than the classical results on additive value functions [KR76], and more practical, too. These theoretical results can be directly utilized in the scope of the developed methodology: First, the user provides us with a set of qualitative preference statements. These statements are used (by means of solving a set of linear constraints) to generate a candidate value function whose structure is based on the qualitative information supplied by the user. The existence of such a value function is guaranteed by the developed representation theorem. The user is presented with the top database items according to this value function and either quits or indicates the best presented item. This information induces additional linear constraints on the value function, resulting in a modified value function and a new set of top candidates. This process continues until the best item is identified.

In collaboration with Ben-Gurion University, the above methodology has been implemented as a part of a prototype system for online flight reservations. Empirical evaluation of this system showed that using the novel methodology the optimal flight configuration is typically identified within a very small number of iterations, and compared these results to the results achieved with a standard technique from multi-attribute value theory. Currently, C. Domshlak is working on extending both theoretical and practical aspects of the developed methodology. This work involves collaboration with researchers from Cornell, Ben-Gurion, and Leipzig universities.

In a Science perspective (2002), Gomes and Selman discuss the complexity of a prototypical hard computational problem, Boolean Satisfiability, that captures the computational core of a range of hard, large-scale computational problems. The perspective also discusses improved algorithms to solve the Satisfiability problem using tools from statistical physics and combinatorics.

Gomes and Shmoys (2002) developed a hybrid approach for combinatorial optimization, consisting of an exact randomized complete search method that tightly couples constraint propagation techniques with information obtained from randomized rounding of linear programming (LP) relaxations. This hybrid strategy outperforms pure constraint programming and LP strategies and other approaches. The hybrid approach relies on the LP relaxation to provide good approximate solutions. For the quasigroup completion problem, a combinatorial problem with structural properties similar to those of a variety of real-world optimization problems, Gomes et al. developed a new approximation that is within a factor of 1-1/e = 0.63 from optimal. The best earlier approximations gave a factor of 0.5. The approximation proposed by Gomes et al. uses randomized LP rounding, based on an LP relaxation of a packing formulation of the problem.

Bejar et al. (2002) have shown tradeoffs between problem representations based on constraint programming, logical representations (Boolean satisfiability encodings), and mathematical programming formulations. Problem representation is a key factor affecting the efficiency of combinatorial search. For example, Gomes et al. have also shown how one can substantially improve the performance of search methods by using primal-dual encodings, by adding inferred (redundant) constraints, as well as by exploiting the well-structured subcomponents of problems with efficient propagation algorithms

Randomized search strategies have been highly successful in local search (e.g., simulated annealing Kirkpatrick83, tabu search (Glover 1989), and genetic algorithms (Holland 1975). However, such methods are inherently incomplete, in that they do not guarantee optimality of the solution. Optimality can be guaranteed using backtrack style methods such as branch-and-bound. These methods can explore the full combinatorial space. Backtrack style search can be randomized by introducing a random element in the variable choice and / or value selection heuristics. With some minimal additional book keeping, one can still maintain completeness of the search strategy. Researchers have observed that the performance of backtrack search methods can vary considerably from instance to instance. There have been some theoretical results showing that randomization can improve the performance of complete search methods (Motwani and Raghavan 95, Luby et al. 93), but randomization was not believed to provide significant practical benefits in a complete search setting. For example, up to about five years ago state-of-the-art Davis-Putnam style Boolean satisfiability solvers did not include randomization. In the work on the study of the run time distribution of complete search methods by Gomes et al., they have demonstrated that one can in fact obtain exponential speedups by randomizing a complete search method. Such speedups can be obtained by taking advantage of the high variance in run time of randomized complete methods. In particular, they have shown that the extreme variance or *unpredictability* in the run time of complete search procedures on combinatorial problems can often be explained by the phenomenon of heavy-tailed distributions. Heavy-tailed distributions are highly non-standard probability distributions, capturing phenomena with infinite moments, for example infinite variance or infinite mean. Previously such distributions have been used to model erratic behavior in, for example, weather patterns, stock market behavior, and time delays on the World Wide Web.

In more recent work, Chen, Gomes, and Selman (2001) developed formal models of backtrack search that provably exhibits heavy-tailed. Athreya (Athreya 2002) formulated the concepts of i) heavy tailed distributions with bounded support and ii) a sequence of asymptotically heavy tailed random variables and application of these concepts to combinatorial search problems. Athreya also showed heavy tailed behavior in combinatorial search on random Galton Watson branching trees using results from classical branching process theory as well his new results on generation of heavy tailed distributions as ratios of independent r.v with positive density at zero are used. In addition to this work, Athreya worked on stationary distributions for Markov chains generated by iteration of iid random maps on R+. The results obtained led to two papers, one on stationary measures for some Markov chains on R+ with applications to ecology and economics (Athreya 2002) and the other on Harris irreducibility of iterates of maps on R+.

The understanding of the extreme variance that characterizes complete search algorithms on combinatorial problems has led to the introduction of novel strategies for the design of algorithms based on "rapid restarts" and "portfolio" strategies (Huberman et al. 97; Gomes et al. 98; and Gomes and Selman 01). In a restart strategy, one repeatedly restarts the search procedure with a new random seed after a certain predefined number of backtracks; in an algorithm portfolio, many copies of a randomized search procedure (each started with a different random seed) or a mix of different search procedures are executed in parallel or interleaved. Restarts and portfolio can significantly reduce the variance in run time and the probability of failure of the search procedures, resulting in more robust and more efficient overall search methods. Gomes and Selman have shown that restarts provably eliminate heavy-tailed behavior. The results of this research have changed the general view of randomization of complete search methods in, for example, the Satisfiability and Constraint satisfaction community. Randomization and restart strategies have now been incorporated into state-of-the-art solvers for the Boolean satisfiability problem (SAT solvers). In fact, the rapid restarts technique is an integral component of the current world's fastest SAT solver, called Chaff (Moskewicz et al. 2001). (In Chaff, restarting is combined with clause learning, another technique central to Chaff's overall effectiveness. In clause learning, certain derived clauses are carried over between restarts.) Chaff can handle problem instances with over one million variables and four million constraints. The solver has been used to verify correctness properties of the latest Alpha chip design, which has a complexity comparable to that of the Pentium IV. Gomes et al. have also demonstrated the effectiveness of randomization and restarts for branch-and-bound search methods, distributed constraint satisfaction, and planning and scheduling problems. Randomized restarts have been demonstrated to be effective for reducing total execution time on a wide variety of problems in scheduling, theorem proving, circuit synthesis, planning, and hardware verification. (Gomes et al. 2000, Gomes and Selman 2001, Moskewicz et al. 2001, Meier et al. 2001). In current research they are investigating optimal policies for restart strategies (Gomes and Selman 2001b, Kautz et al. 2002, Horvitz et al. 2001).

Fernandez et al. (2002) are studying Distributed Constraint Satisfaction Problem (DisCSP) formulations, which allow to model combinatorial problems arising in distributed, multi-agent environments. Their work focuses on the following topics:

- Study of the performance of some DisCSP algorithms on real communication networks
- Identification and formulation of real world problems as DisCSP
- Development of applications for wireless devices running DisCSP algorithms

Fernandez et al. proposed a benchmark for the evaluation of distributed algorithms, SensorDCSP. Sensor DCSP is a naturally distributed benchmark based on a real-world application that arises in the context of networked distributed systems. It is also inspired on a DARPA challenge problem proposed in the context of the ANTS initiative.

In order to study the performance of DisCSP algorithms in a truly distributed setting, Fernandez et al. use a discrete-event network simulator, which allows them to model the impact of different network traffic conditions on the performance of the algorithms. They considered two complete DisCSP algorithms: asynchronous backtracking (ABT) and asynchronous weak commitment search (AWC). In their study of different network traffic distributions, they found that random delays, in some cases combined with a dynamic decentralized restart strategy, can improve the performance of DisCSP algorithms. More interestingly, they also found that the active introduction of message delays by agents can improve performance and robustness, while reducing the overall network load. Finally, their work confirms that AWC performs better than ABT on satisfiable instances. However, on unsatisfiable instances, the performance of AWC is considerably worse than ABT.

Shoemaker and Regis are studying global optimization through machine learning and response surface methods The objective of their study is to use a combination of heuristic search algorithms, nonlinear response surface methods, and machine learning techniques to more efficiently find solutions to nonconvex optimization problems in which evaluation of the objective function is computationally costly. Such techniques have enormous potential in engineering where each objective function evaluation requires lengthy simulation of complex computer codes. Optimization for these kinds of problems is important both in system design to maximize performance and in parameter estimation (the inverse problem). One important class of especially costly functions are those that require solutions of a system of partial differential equations, where nonlinearities and accuracy requirements necessitate small time steps and fine meshes, resulting in large computation times for each model simulation. These efforts have focused on the three following areas:

1.1 Learning Response Surfaces

There are shortcomings with most of the existing methods for optimization of computationally expensive functions. Gradient-based methods cannot be used sometimes because derivatives are too expensive or impossible to compute. Evolutionary algorithms and other heuristics like simulated annealing often require an enormous number of function evaluations to obtain adequately good solutions. An alternative is to learn and predict trends in the underlying objective function based on a limited number of objective function evaluations (data points). They are working on papers in this area including one on Parallel Radial Basis function Methods for Costly Nonconvex Optimization.

1.2 Enhancing Evolutionary Algorithms with Response Surfaces

As mentioned earlier, evolutionary algorithms often require a very large number of function evaluations in order to obtain an adequately good solution. By using response surface techniques in combination with an evolutionary algorithm, it is possible to substantially improve the performance of the evolutionary algorithm for computationally expensive functions. The response surface model is used to screen out less suitable offspring and to limit the costly function evaluations to those offspring that appear most promising based on the current response surface. They are doing the calculations for combining a global radial basis functions with evolutionary algorithms and will eventually write a paper in this area.

In addition, they are finishing a manuscript on evolutionary search methods used in conjunction with both quadratic and local radial basis function approximation (based on nearest neighbors) that is expected to be submitted within a week. The radial basis function is coupled with a polynomial for the response surface and a symmetric Latin Hypercube experimental design is used to select initial objective function values for evaluation. In this paper they found that our local radial basis function approaches worked much better on higher dimensional problems than quadratic approximations on the higher dimensional test functions and better than the evolutionary algorithm without a local approximation. On lower dimensional problems, both the local radial basis and quadratic response surfaces performed better than the conventional evolutionary algorithm without a local approximation.

1.3 Application to Difficult Real Problems
They are applying some of the algorithms discussed above to systems of partial differential equations arising in engineering. The particular application is a system of partial differential equations describing the movement on contaminants in groundwater and the effect of injections on biodegradation of the contaminant. One of the major areas where such contamination occurs is at military bases, including airfields and the military has had to commit a significant portion of its budget to contamination removal. The goal of the optimization is to find the least expensive way to detoxify the site.

Theme 3 – Pervasive computing: autonomous distributed agents and communication networks

Under this theme we have five main projects:

1) Control of Autonomous and Semi-Autonomous Vehicles (D'Andrea)

In order to fully exploit the new capabilities offered by faster computation, advanced sensors and actuators, and technology in general, control theoreticians must bring to bear new tools and techniques used in other disciplines. For example, there is potentially great synergy between computer scientists and control theoreticians for tackling new problems in distributed and hierarchical control of autonomous systems. Applications include any "dull, dirty, or dangerous" situation where autonomous entities aid human beings, such as space exploration, disaster relief, and national defense.

One of the long-term research objectives of D'Andrea's research group is to understand how the problems typically studied by computer scientists, such as artificial intelligence, distributed computation, and program verification, can augment control research in these application areas. In their research project, Control of Autonomous and SemiAutonomous Vehicles they have completed a new, comprehensive multi-vehicle testbed, which is being used for research in control of multi-vehicle systems. The test-bed consists of 24 ground vehicles that engage in a game of "Capture the Flag", with and without limited human control. A fully simulated version of the test-bed has also been developed, and is being distributed to other researchers interested in these types of problems. Various research groups attended a group competition on September 23-September 27, 2002.

A Mixed Logic Dynamical (MLD) framework has been used for coordinating strategies of these vehicles. The approach is computationally intensive, but does lead to nearly optimal strategies. A probabilistic approach for vehicle path planning in uncertain environments has been developed and tested on a simulated environment. A joint Caltech/Cornell summer program was successfully completed on August 23. The program consisted of three Cornell students spending the summer at Caltech, and three Caltech students spending the summer at Cornell, with the objective of engaging in a series of games of "Capture the Flag" on the multi-vehicle test-bed at the end of the summer. The results of this summer long research program will be presented at an invited session (6 papers) at the American Control Conference in 2003. The Cornell RoboCup team competed in the World Robot Soccer Championships in Fukuoka, Japan. The team beat Germany in the final 7-3. The Cornell team has won this competition in three of the four years it has competed, placing third the only time it did not win.

2) Activity Compass (Kautz)

The Activity Compass project is a wireless PDA-based application that provides active assistance in achieving transportation goals. The system learns to predict the user's destination goal based on previous behavior and current context, rather than explicit programming. Further, the system reasons about the probability of success in achieving the goal in relationship to a real-time data feed of transportation conditions. If the system determines a likelihood of failure, it actively prompts the user to modify his current plan to one that has a higher success probability.

The information that the Activity Compass receives from the user consists of a series of Global Position System (GPS) readings that contain current position, and velocity information. The user is expected to have carried the Activity Compass for a period of weeks prior to expecting the Activity Compass to add any value to the hand-held computer, and so a historical database of GPS readings are expected to exist as well. The user interface is minimalistic: the Activity Compass outputs an arrow and an icon, which instructs the user on how to reach a destination.

The basic model of a user that the Activity Compass maintains is called an Activity Path. An Activity Path is an abstraction of the sensor readings that are generated by user behavior. For example, the Activity Compass might maintain an Activity Path that corresponds to finding your parked car after work. Generating Activity Paths requires segmentation of the input data stream into semantically coherent pieces, and abstracting relevant details from multiple corresponding segments of data. The server, while off-line, periodically does these computations. This processing corresponds to training the device and is only done incrementally to incorporate new knowledge into the Activity Compass. Activity Paths capture relationships between time, user location and mode of transportation, and can be partially abstracted to capture concepts such as "home", "bus stop", or "morning."

When active, the Activity Compass monitors which Activity Paths it believes are in progress at the current time, and what constraints are in place that might prevent Activity Paths completion. Determining constraints involves integrating a calendar, real-time bus and traffic information, perceived user-preferences and knowledge about the transportation domain. Once constraints have been identified, the Activity Compass can choose a destination that satisfies the most important constraints, and direct a user toward it. For example, the Activity Compass might reason that in order to get home, a user must take a bus, and so it is currently more important to direct the user to a bus stop than it is to direct a user home.

The Activity Compass is currently implemented as a client/server architecture. A Palm i705 hand-held computer functions as the client and a 1.5Ghz Pentium II networked computer running the Linux operating system functions as the server. Despite the abundance of consumer grade hand-held computers, GPS receivers, cellular telephones, and wireless devices, there are very few options for devices that combine all of these technologies. The convergence of these devices, coupled with a suitable battery life, is a prerequisite for a consumer-grade Activity Compass. In the meantime, we have developed a custom-built battery pack and GPS receiver. Together with the handheld computer the entire device is approximately the size of a large novel and only lacks detailed location accuracy (< 15m) and telephone capabilities. The Activity Compass utilizes a very low bandwidth cellular telephone network connection to communicate current sensor readings to a server. The server responds with a user interface task that the client carries out. By utilizing the local GPS sensor and the computational ability of the client, the system is moderately robust to temporary communication failures with the server.

In addition to the hardware, they have developed the overall system architecture, the data gathering module, communication routines, and an initial user interface. In the next work period they will implement the learning and intervention algorithms.

3) Wireless Application Development (Fernandez)

Rapid growth and spread of wireless communication devices has lead to developers to adopt efficient solutions in terms of software development. A good alternative is the use of platform independent languages, such as Java, that help developers to write applications unaware of device specific characteristics. Fernandez et al. have focused on the development of tools to help new developers to establish a programming environment for wireless devices, such as PDA, for a multi-platform scenario. They have also developed distributed wireless applications.

More specifically, they have adapted and modified a programming platform for small devices, known as Waba. Waba's language, virtual machine and class file is ideal for small devices, being a strict subset of the syntax of the Java language. Fernandez et al. improvements over Waba allow developers to use Java Remote Method Invocation in order to access manufacturer device drivers for communications technologies such as Bluetooth devices. In addition, they have also developed some graphical applications for Bluetooth Networked PDA devices in order to make easier a deployment of future implementations of Distributed Optimization Algorithms.

In relation with the work mentioned above on DisCSP, Fernandez et al.'s ongoing projects aim at the following directions:

- To identify useful scenarios for Distributed Optimization Algorithms.
- To develop and link such algorithms under our wireless development platforms.
- To study and implement routing algorithms for Bluetooth networks.

Fernandez et al. work is directly related and supports Spina's research group.

4) Reasoning about knowledge and uncertainty in multi-agent systems (Halpern)

Halpern's research is concerned with representing and reasoning about knowledge and uncertainty in multi-agent systems. The work uses tools from logic (particularly modal logic and the idea of possible-worlds semantics), probability theory, distributed systems, game theory, and AI, and he likes to think that it contributes to our understanding of each of these areas as well.

Some themes of his current research include: (1) applying ideas of decision theory to constructing algorithms in asynchronous distributed systems, database systems, and wireless systems, (2) providing foundations for useful qualitative notions of decision theory, (3) reasoning about security.

5) Human Computer Interaction (Sengers)

The goal of Sengers's research is to address problems in Artificial Intelligence and human-computer interaction that bridge cultural issues and technology design. In order to do this, Sengers uses cultural analysis to evaluate the cultural meaning of current systems, and then develop new systems which express different meanings. She is particularly focused on consumer culture and everyday computing, or designing technologies for everyday life which have a positive cultural impact, and on analyzing and supporting reflection on emotional experiences.

Sengers is engaged in a coalition of researchers from Cornell, Georgia Tech, Royal College of Art, Swedish Institute of Computer Science, and Intel called Affective Presence which explores how to design ubiquitous computing devices that reflect and engage human emotional, social, and spiritual experiences. She has completed work on the EU SAFIRA project, which developed applications in and a toolkit for affective computing, or computing that responds to human emotional states. She is also involved of several major projects exploring affective computing, including the Fear Reflector, a device to support reflection on fear while backpacking; the Home Health Monitor, a networked 'smart' system for the home which supports reflection on a household's emotional climate by generating 'home horoscopes', and the Affector, a device to help friends in neighboring offices get an ambient sense of each other's moods.

In terms of new findings, Sengers's team have run laboratory experiments using physiological sensing under conditions of physical exertion and have discovered that many simple laboratory measures of fear are not applicable to this more physically strenuous situation. They are developing a new approach together with Thorsten Joachims using statistical learning techniques to uncover broader patterns in multiple simultaneous channels of physiological data. They have collected a database of 4,000 horoscope sentences and have developed semi-automatic techniques for generation of plausible horoscopes which touch on emotional issues. They have developed a system to classify the sentences with user feedback according to their emotional content. They have developed an early prototype of the Affector system which uses distortion of video signals to communicate a subjective sense of mood or emotional climate.

Theme 4 – Advanced Computing Architectures

Continuing improvements in device fabrication, VLSI design, and packaging have ushered in an era of ubiquitous computing. It is now possible to embed processing capability for extremely low cost in places previously not imagined. In traditional computer systems, data is moved around the system to the processors, memories, or devices that need it.

The research in advanced computing architectures at IISI mainly involves two main projects: Active Memory Clusters, and the study of the effects of thread migration in MPI runtime layers through the development of the Tern runtime system.

1) Active Memory Clusters (Heinrich)

The Active Memory Clusters project seeks to provide the performance of hardware distributed shared memory machines at a fraction of the cost of custom machines available from hardware vendors. Active Memory Clusters utilizes new techniques developed in memory controller design to allow the creation of low-cost shared memory machines utilizing commodity clusters of high-end servers. Heinrich et al. are currently investigating the performance benefits of combining active memory techniques with multiprocessor systems in order to further improve performance Henrich's research through IISI focuses on data-intensive computing, the ability to embed processing capability in the memory and I/O subsystems of traditional computer systems, dramatically improving their performance. The key research insight is that active memory support in the memory controller can be viewed as an extension of the cache coherence protocol if the memory controller is designed in an exible manner.

They are now extending their novel active memory system design to include a two-level approach to active memory systems that focuses on an active memory controller that leverages the cache coherence protocol to allow application address re-mapping techniques to improve cache behavior, and active memory elements that can assist an active controller in performing data-intensive operations in the memory system itself. They also have ongoing research in novel active memory applications and address re-mapping techniques, especially in multiprocessor active memory systems. They continue their active memory clusters design that merges the research ideas of hardware distributed shared memory, active memory systems, and clusters. Industry's InfiniBand and 3GIO networks represents an enabling technology for this research, as their node integration point combined with their novel memory controller will allow hardware DSM machines to be constructed from industry-standard workstation clusters. They plan to prototype and experiment with variations of their AMC memory controller via an FPGA implementation.

Heinrich et al. showed that address re-mapping techniques to improve cache behavior can be implemented transparently by leveraging and extending the cache coherence protocol via an exible memory controller. Such an active memory controller can provide up to 7x performance improvement on uniprocessors. They also demonstrated the first use of active memory address re-mapping techniques on both single-node and multi-node multiprocessors, as well as the novel coherence protocol extensions that support them. They designed a new memory controller architecture that merges the research ideas of hardware DSM, active memory systems, and clusters. They call this architecture Active Memory Clusters (AMC). The resulting architecture maintains the cost advantages of current clusters while attaining the performance of hardware-based shared memory machines.

Unlike other similar proposals, the active memory clusters work benefits from the fact that the enhanced active memory controllers used in AMC also improve the performance of single -node systems, a requirement for inclusion in any commodity machine produced by the likes of Dell, Compaq, HP, or Intel. They have been able to show that such a system can significantly improve the performance of both uniprocessor nodes, and clusters of active memory-enabled machines (see Chaudhuri et al. 2002, Kim et al. 2002, and Heinrich et al. 2002).

2) Study of the effects of thread migration in MPI runtime layers through the development of the Tern runtime system (Speight)

The Tern project is a result of the increasingly difficult of achieving proper load balance and concomitant performance on large-scale message applications for clusters of high-end servers. The Tern project has resulted in 1 conference paper currently outstanding for publication, 1 journal paper in preparation, and the Master's Thesis for Jian Ke. Speight et al. have developed a set of extensions for the MPI specification that give the programmer the ability to install custom thread migration algorithms, or to use one of the "canned" algorithms provided with the Tern system. They are currently working with collaborators in the Cornell Theory Center to examine the issues related to scaling Tern from their own cluster of 16 processors to the 256 processor Velocity cluster available from the Theory Center. The Tern system has shown that providing thread migration in MPI runtime layers can results in performance improvements of up to 2x for applications with irregular communication patterns. They are continuing to develop thread migration policies, investigating the effect of thread migration in heterogeneous environments, and providing multiprogramming support in the Tern runtime system (see Ke and Speight 2002).

3. Research Accomplishments – AFRL Researchers

Opportunistic Behavior in Multi-Agent Planning and Execution

(James Lawton and Carmel Domshlak)

Single-agent opportunism is the ability of an agent to alter a pre-planned course of action to pursue a different goal, based upon a change in the environment or in the agent's internal state -- an opportunity [Ham93,Law03]. In their project, James Lawton and Carmel Domshlak extended this notion to multi-agent opportunism - the ability of agents operating in a multi-agent system (MAS) to assist one another by recognizing and responding to potential opportunities for each other's goals [LD03a,LD03b,LD04a]. In theory, multi-agent systems can clearly benefit from the ability of agents to act opportunistically. In practice, however, taking advantage of an opportunity at an inter-agent level is far from trivial: The agents should be capable of recognizing whether a given event or situation may be an opportunity for a goal of another agent in the system, and of responding appropriately to these recognized opportunities. Two key issues can make the potential practical attractiveness of multi-agent opportunism somewhat questionable. First, both recognizing opportunities and responding to them should have low computational complexity, otherwise the MAS will be more ``socially friendly" than useful. Second, in order for an agent to recognize potential opportunities for other agents, the agent clearly has to know something about what these other agents are doing. Unfortunately, in many applications this ``something" tends to be very limited. The question is whether multi-agent opportunism can be effective in such cases of extremely limited shared knowledge?

In their work, James Lawton and Carmel Domshlak considered multi-agent opportunism in systems where the agents are required to perform non-trivial planning tasks. The planning and execution scheme for the agents used in their study was developed especially to support opportunistic behavior of agents in various settings of shared knowledge. This scheme is based on the machinery developed on top of partial-order plan graphs (or POPGs, for short) [LD03a], extended with predictive encoding of opportunities [Pat91] (in short, potential opportunities are pre-computed and associated with existing plan elements). In their empirical study, James Lawton and Carmel Domshlak focused on the somewhat ``least permitting" conditions of shared knowledge, and examined whether multi-agent opportunistic behavior can be beneficial in this settings. First, they assumed that no online re-planning is allowed/possible. Second, they considered two (probably the most basic) settings of shared knowledge. In both settings, the agents communicate only information about their suspended goals, i.e., goals that they can no longer achieve. In addition, the agents were assumed to have a priori only very limited knowledge about the other agents in the group: In the less informative settings, the agents know only the ``types" of the other agents in the MAS, i.e., their individual capabilities. In the more informative settings, the agents know about the individual goals that have been assigned to the other agents.

The variant of predictive encoding considered by James Lawton and Carmel Domshlak in their project is based on the idea of planning for capabilities. In short, each agent plans not only for its assigned goals, but also for a limited subset of its other, opportunity-wise ``most promising", capabilities. What makes certain capabilities more promising than others, and whether such a ranking of capabilities can be any more effective in the face of limited shared knowledge, were the main questions that James Lawton and Carmel Domshlak studied in their research. For the evaluation, they implemented a discrete-event simulation platform for MASs using the POPG-based planning and execution scheme. Using this platform, they have conducted a set of experiments corresponding to several levels of mutual knowledge shared by the agents in the MAS. The benchmark problems used in the evaluation were based on the standard planning benchmark domain inspired by the planning problems for NASA's Mars Rovers, used in International Planning Competitions.

The results of the evaluation performed by James Lawton and Carmel Domshlak show that adopting opportunistic behavior for planning agents does not come for free, and that its efficiency depends significantly on the way the shared knowledge is exploited by the agents. On the positive side, the results show that multi-agent opportunism can improve the overall performance of an MAS, even in extreme situations where the amount and type of the shared knowledge are very limited, and when the agents have little or no ability to re-plan. However, the basic planning and execution mechanism that has been developed to achieve multi-agent opportunism often produced inefficient plans, occasionally resulting in reduced system performance. Therefore, in their recent work, James Lawton and Carmel Domshlak have introduced a set of extensions to their original approach to multi-agent opportunistic planning and execution aimed at improving the efficiency of the plans and the system performance [LD04b]. In particular, they have examined two post-planning methods of enriching the core structure of a plan: one that adds ``shortcuts" bypassing the segments of the plan devoted strictly to support predictively encoded extra goals, and one that predictively repairs the core plan to include subplans achieving the extra goals. In addition, they have examined an online approach that assumes the agents posses limited runtime plan repair capabilities. Using this approach, the agent attempts to enhance its core plan only at the time it learns of a goal suspended by another agent. The results of the empirical evaluation conducted by James Lawton and Carmel Domshlak demonstrated that when some simple measures are taken to augment the core plans, multi-agent opportunism is indeed feasible in that it produces results as least as good as, and often better than, not using multi-agent opportunism. Further, this improvement can be obtained even when the agents have only very limited knowledge of each other's capabilities, and even when the agents have no ability to re-plan at runtime.

This project is part of the dissertation of James Lawton (AFRL/IFTB), who is pursuing a PhD at the University of New Hampshire. Lawton's dissertation advisors are Professors Elise Turner and Roy Turner at the University of Maine. Carla P. Gomes (Cornell University / IISI) is also providing research guidance to Lawton through the IISI, as well as serving on his dissertation committee.

Adaptive Information Transformations in JBI
(Carmel Domshlak and Mark Linderman)

The Joint Battlespace Infosphere (JBI) is a combat information management system that provides individual users with the specific information required for their functional responsibilities during crisis or conflict [JBI99]. A JBI is created by the joint task force (JTF) commander, and is managed at this level by the information management stuff. The general purpose of JBI is to integrate data from variety of sources, aggregate this information, and distribute it in the *appropriate form and level of detail* to users at all echelons. In particular, while building a global view on the current situation, the JBI *tailors this picture for individual users*: the commander gets a high-level view of the campaign, while the soldier in the field gets a detailed description of a nearby hostile base. Each warfighter receives from JBI exactly the information needed to perform his or her function, and this information is continuously updated with the new relevant data entering JBI. In short, the information processing in JBI is based on two key concepts: (i) Information exchange through "publish and subscribe", and (ii) Transforming data into knowledge via fuselets. Fuselets are programs that act as clients internal to JBI [JBIF03]. When information becomes available to a fuselet via a subscription, the fuselet publishes one or more new objects as a result. The conceptual task of fuselets is to process the incoming data with respect to commander's standing orders, general needs of JTF, etc. Technically, *fuselets are scripts that are implementing some standard/specialized sets of rule.*

Acquiring and presenting information are knowledge-intensive activities: The standard task of a JBI fuselet is to collect a coherent body of information into a document, structure and transform it in a way meaningful to a certain set of clients. The problem is that the needs of the JBI clients from the same set of data can be very different. Therefore, the number of alternative *views* on the data that should be created by the fuselets ahead of time is potentially extremely high, and this may harm the scalability of the JBI platform. Addressing this problem of pre-compiled views, Carmel Domshlak has been developing a general framework for fuselets' artifacts, based on annotating information objects with a structured set of *conditional standing transformations*, reasoning about which is performed using novel models and techniques developed in the field of AI [BDS01,BDS04]. The key idea behind this framework is to allow data experts (i.e., JTF personal) to specify in an intuitive, compact, yet expressive manner how the data should be presented to the clients conditioned on their type, status, history of actions, etc. To make the whole framework practically appealing, the language of such specification is kept purely *qualitative*, consisting of statements of constraints, preference, relative importance, etc. [BBDHP03a,BBDHP03b]. These standing transformations are encoded as part of the object's meta-data, and used to determine the actual content and form of the information object when this is actually requested by a concrete client.

The whole framework of adaptive information objects developed by Carmel Domshlak (together with his colleagues from Ben-Gurion Univ. and Tel-Aviv Univ.) has been already implemented as a part of three different prototype information systems [BDS04]. Currently, Carmel Domshlak is collaborating with Mark Linderman from the JBI team in AFRL on further specification of this framework in lines of the JBI infrastructure.

Mixed Initiative Decision Making
(Joe Carozzoni, C. Domshlak, Carla Gomes, and Jerry Dussault)

Classical decision theory assumes that (i) the set of possible decisions and the probability distribution over outcomes is known in advance, (ii) the utility an agent assigns to any possible outcome is fully specified in advance, (iii) the goal of each agent is to maximize his or her expected utility, and (iv) the computational cost of finding a solution need not be considered. Real-world strategic decision making situations rarely meet all these requirements. The initial decision problem is typically ill defined, and the model of the decision problem grows through many iterations as flaws and inconsistencies are revealed. The user's utilities are not always known in advance, but may only be determined incrementally as he accepts or rejects candidate solutions. Furthermore, a useful solution to a problem is not just a decision, but rather a defensible reason for making that decision, in terms of the facts and assumptions built into the model. Finally, both the human and machine effort needed to make the decision must be taken into account according to the broader decision context, which can range from long-range planning to immediate action under fire.

Research in Artificial Intelligence deals with many of these issues through work on interactive planning, preference elicitation, resource-bounded reasoning, and algorithms for single and multi-agent decision problems. Some researchers in psychology and decision science (such as work in judgment under uncertainty and naturalistic decision making) also address situations with poorly defined goals, missing data, stress, high stakes, time pressure, and uncertainty. Studies in fields such as military command and control show that the strategies human experts employ are quite distinct from the simple classical model. For example, classifying a situation as an instance of a previously solved problem is a more prevalent strategy than systematic weighing of alternatives.

Following this motivation, the IISI/Cornell researchers have collaborated with AFRL/IFS on defining opportunities for research on critical decision making. The key event of this continuous collaboration was the AFRL/IISI Workshop on Mixed Initiative Decision Making that brought 35 computer scientists, psychologists, decision making analysts, and military personnel to the Intelligent Information Systems Institute at Cornell University to perform a gap analysis for the areas of human decision making, preferences, time criticality and uncertainty, and multi-agent systems. The presentations and discussion at the workshop have been extremely fruitful, and we believe that the workshop successfully identified key research opportunities to develop technology that will be broadly useful for DoD decision support applications. The detailed report of the workshop is available at [MIDM03].

Nate Gemelli
Mining Social Networks

Gemelli worked with Hopcroft and Selman on discovering hidden structure from large repositories of data. Two Cornell grads, Omar Khan and Brian Kulis, also took part in this research project. Discovering patterns or sub-graphs (communities, isomorphic graphs, etc.) from a larger dataset/graph is an important aspect of identifying important formations of information sets (i.e. knowledge) in large amounts of data that would otherwise be overlooked by a human observer. The purpose of this research was to be able to identify unique isomorphic sub-graphs within domain specific data represented as a large input graph. Graph pattern matching techniques were used to identify these unique isomorphic sub-graphs, and an output set S, which contains these sub-graphs, is presented to a user for further analysis.

Gemelli also worked with Spina on the Information Routing in a Dynamic Wireless Network project (see below).

Capabilities Aware Routing (CAR) is another project Gemelli was focusing on. CAR is a project that originated out of work from John Spina's previous research in Ad-hoc Wireless Networks. The fundamental objective of this project was to be able to route information intelligently from one node in an ad-hoc wireless network to another node in that same network taking into consideration the properties of each node that the route path will traverse. In doing so, they envisioned that information can be delivered to its destination in a quasi-optimal way. For instance, if they are dealing with a Bluetooth (BT) wireless network, where there is low bandwidth, low power, and limited range (sub 300 feet) of the communication, and they need to get a file to a laptop 500 feet away (well out of range of level-2 type BT devices), then there is a capability that cannot be fulfilled by BT type technology. However, there exists a device in their BT network that is both a BT enabled device as well as an 802.11 enabled device (much farther range of communication). Using CAR, they will be able to recognize the limit of BT, search for higher powered means of communication, such as their 802.11 device that is in their network, and hand off information to be transmitted (from BT to 802.11). They will enable a seemingly helpless network (helpless with regards to range) with the power to communicate its resources (i.e. information) outside its otherwise limited bounds.

Nate Gemelli is also working on the Adversarial Agent Environments (for Wargaming)
The purpose of this research is to take a real world scenario, model it as a graph, and put adversarial agents within the environment to complete some given tasks. By defining their environment as a graph, nodes representing areas occupied by an agent and edges being a logical connection between areas, they can model real world scenarios, such as urban warfare. This research falls in the class of Transport and Search and Rescue problems. Assume that they have two teams of agents, Red and Blue. Red's objective is to protect some resource that is within the environment, and to eliminate Blue from its environment (either by destruction or making Blue's agents flee). Blue's objective is to

find the resource they were ordered to find (i.e. prisoners, intelligence reports, other previous Blue team members who may be pinned down), and extract that resource out of the environment. The interesting part about the Blue team is that the team may not be entirely composed of cooperative agents. Blue team may represent a coalition force (e.g. US and Great Britain). Of course, global utility will state that the goal must be completed by the Blue team in general, with no regards to which part of the coalition completed the mission. Internal to each agent, there is a utility which will also reward for a completed mission, however, that internal utility will reward higher if the mission is completed (i.e. resource extracted) by their side of the coalition forces. They will use non-cooperative game theory to model the interactions of coalition forces. This is new research and formal models have yet to be defined.

Robert Paragi

This work focused on the theme of studying the effect of structure on problem complexity using graphical tools.

The objective was to provide a better human interface to the problem structure and the ongoing solution process.

The environment for this work was the quasigroup completion problem, also known as a Latin square. A quasigroup is a matrix of individual cells in a two-dimensional grid, colored so that each cell receives a color, which is not repeated anywhere else in the given cell's row or column within the matrix.

The color assignment problem was represented by logic expressions for the constraints on the problem, such as those limiting a color's use only once in a given row and column. Local search constraint solving was the method used to complete the assignment of colors to all cells in the matrix. One or more solvers were available from existing tools whose construction and operational algorithm improvement were outside the scope of this effort.

Emphasis was put on making a Java graphics display (of incremental progress toward quasigroup completion) communicate consistently with a constraint solver.

The Java display program and the C constraint solver program are two processes communicating across two sockets. Across one socket the Java program is a client of the C program's command server: the user enters a command from the Java program's GUI, which acts in the role of a command-processing client in this situation, and the command is then sent to the C solver, which acts as a command-processing server. Client-server roles are reversed between the programs for data transmission across the other socket. In this situation the C solver is the data-display client, and the Java program's GUI is the data-display server. The Java program had already been a multi-threaded program when acting only as a data-display server. After a second socket was added (on a port different from that used for data transmissions) for communication in command entry and processing functions, inconsistent behavior became the focal point. In the C solver a separate thread was later added solely for socket communication related to commands only, and in situations where it did function for a while, communication of commands seemed to have functioned well for longer periods of time than before it was made a separate thread in the C solver, but more testing later showed that correct operation could not be consistently maintained. Experiments with raising and lowering thread priorities have not helped. Work continues on this problem.

Carla P. Gomes and Ramon Bejar (Cornell University / IISI) provided research guidance to Paragi through the IISI.

Nancy Roberts

Bayesian Predictive Model of an Interactive Environment

The objective of this project was to apply uncertainty techniques (Bayesian Networks and Decision Theory) to COTS tools in the area of home automation and thus, add intelligence to it. Roberts worked to finish an AFRL Inhouse Technical Report as part of this project. Carla P. Gomes and Mike Pitarelli (Cornell University / SUNY IT) provided research guidance to Roberts through the IISI.

In collaboration with several researchers at Cornell, Nancy started a project on AFRL 3D World and Intelligent Avatars Project. The objective of this project is to explore and apply various artificial intelligence techniques to enhance a digital informational environment. The environment in this case is a 3-D world based on Active Worlds™ environment used to provide information about AFRL. The AFRL 3-D world initially will focus on creating an informational world that targets the general public and will focus on learning more about the lab, our history, the technology we work on, and the Air Force. Throughout this world, intelligent avatars (or bots) will provide assistance to visitors to this world.

John Spina

Information Routing in a Dynamic Wireless Network

Description of project: The main objective of this research effort was the routing of information in a dynamic wireless network using hand-held devices. The long-term goal was to develop intelligent technology that would enable the delivery information to the dynamic user independent of the user's computing device. Based on the user's dynamic, "on the fly" profile, intelligent reasoning would determine the most efficient way of pushing the information to the user. The user profile would contain such parameters as computing device, memory, processing power, screen-size, bandwidth and communication reliability. By exploiting the "built-in" intelligence, the user would get the right information at the right time based on these parameters. The scope of this research effort focused on developing a short range wireless network that demonstrated information routing using hand-held computing devices. The objective was to enable more efficient node discovery, or getting nodes in the network to communicate with each other and pass information within this small network. Once accomplished, the next goal was alternate route discovery, or finding the "best" route from source node to destination node.

The idea of routing information in a wireless dynamic network using hand-held devices was interesting and had practical application to the commercial world as well as the military war-fighter. This effort had particular relevance and application to the Air Force's Joint Battlespace Infosphere (JBI) program.

There had been continued cooperation between AFRL/IF and Cornell University researchers. The team included: Ramon Bejar and Cesar Fernandez of Cornell University, John Spina, Nathaniel Gemelli, and Robert Mineo of the AFRL Information Directorate. The team worked in parallel to develop small application programs that demonstrated node discovery within a Bluetooth piconet and the routing protocols to efficiently route information outside the boundaries of the piconet. They also developed a collaboration program that created an environment that allowed for the exchange of text messages and files among devices within a Bluetooth piconet. They investigated a peer-to-peer protocol called JXTA, developed by SUN Microsystems, as a method for the further development applications that allow connected devices to communicate and collaborate as peers.

Several briefings were presented to both Air Force and Cornell University groups. John Spina wrote a report that satisfied the requirements of a graduate level course at the SUNY Institute of Technology.

Matthew Thomas

Probabilistic Target Tracking with a Network of Distributed Sensor Agents

The purpose of this project was to study the behavior of a network of distributed, intelligent agents, which use probabilistic reasoning to track targets. The probabilistic reasoning used by the agents is in the form of a dynamic decision network that each agent uses to select its best, next action given its beliefs about state of the world. A dynamic decision network is a probabilistic method for determining the most favorable action in dynamic environments and was first outlined in [DW91]. The inference method proposed by Gilks and Berzuini for dynamic Bayesian models, which these agents use, is an improvement on the original algorithms used in such networks because it avoids the progressive degeneration that's symptomatic of these earlier inference methods [GB01]. The beliefs used to make these decisions are based on information collected by each agent with its sensors and are in the form what the agent sees, what it hears and how it feels. What the agent can see is a target if the target is within range of its eyes, what it hears are messages from its neighbors that are within range of its ears, and how it feels is related to its current energy level. Given its belief about the state of the world, the agent has to decide whether is has enough energy to look and/or listen or whether it should rest and conserve energy. A number of such agents are tied together into a network through communication channels and a common task: tracking targets. These agents are distributed in the sense that they are spatially separated and don't share a common, centralized reasoning that controls their behavior. These agents are also considered intelligent because they act autonomously, are reactive and proactive, and participate in some form of social interaction [Wool99]. By studying the interaction of these agents, he is working on creating a method of controlling the behavior of intelligent agents in such a way that ensures that they act cooperatively to accomplish a common goal while reasoning and acting independently.

Thomas is currently developing a simulation environment using SWARM. SWARM is collection of programming libraries, which were originally created for the object-oriented language Objective-C but have since been implemented for JAVA and other languages. These libraries offer a number of functionalities that make creation of multi-agent experiments simpler and make replication of these experiments easier. Thomas' goal is to use the results of these experiments as the basis of his master's thesis, which he will present and defend at Syracuse University. Carmel Domshlak (Cornell University IISI) provides research guidance to Thomas through the IISI.

Louis Pochet

Active Memory Tradeoffs

The goals of the project were:

> Understand architectural limitations on performance of Air Force applications
> Profile target applications wrt to memory access patterns
> Develop new architectures that mitigate the memory bottleneck for those applications

Louis Pochet was part of the research team of a Cornell project on Active Memory Systems led by Mark Heinrich (see above) and Evan Speight (see above).

(Louis Pochet is no longer with AFRL/IF.)

References

[BBDHP04a]
C. Boutilier, R. Brafman, C. Domshlak, D. Poole, and H. Hoos, CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements, *Journal of Artificial Intelligence Research*, 21:135-191, 2004.

[BBDHP04b]
C. Boutilier, R. Brafman, C. Domshlak, D. Poole, and H. Hoos, Preference-based Constraint Optimization with CP-nets, *Computational Intelligence*, Special Issue on Preferences in AI and CP, 20(2):137-157, 2004.

[BDK04]
R. Brafman, C. Domshlak and T. Kogan, On Value-Function Representation of Qualitative Preferences, *Proceedings of the Twentieth Annual Conference on Uncertainty in Artificial Intelligence*, 2004.

[BDS01]
C. Domshlak, R. Brafman and S. E. Shimony, Preference-based Configuration of Web Page Content, Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, 1451-1456, 2001.

[BDS04]
C. Domshlak, R. Brafman and S. E. Shimony, Qualitative Decision Making in Adaptive Presentation of Structured Information, A*CM Transactions on Information Systems*, 2004, in print.

[DW91]
T. L. Dean and M. P. Wellman, *Planning and Control*, Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[DL04]
C. Domshlak and P. La Mura, Game-theoretic entropy, *Proceedings of the Sixth Conference on Logic and the Foundations of Game and Decision Theory*, 2004.

[GB01]
W. R. Gilks and C. Berzuini, Following a moving target--Monte Carlo inference for dynamic Bayesian models, *J. R. Statist. Soc. B* (2001) 63, Part 1, pp. 127-146.

[Ham93]
K. Hammond, Opportunistic Memory, *The Journal of Machine Learning*, 10(3), 1993.

[JBI99]
Report on Building the Joint Battlespace Infosphere, Volume 1: Summary, United States Air Force Scientific Advisory Board SAB-TR-99-02, December, 1999, cleared for publication in February 2000.

[JBIF03]
J. Milligan and J. Hendler, JBI Fuselet Definition Document, http://www.rl.af.mil/programs/jbi/documents/fuselet-definition.doc, September 2003.

[KR76]
R. L. Keeney and H. Raiffa, *Decision with Multiple Objectives: Preferences and Value Tradeoffs,* Wiley, 1976.

[Lang02]
J. Lang, From Preference Representation to Combinatorial Vote, *Proceedings of the Eight International Conference on Principles of Knowledge Representation and Reasoning*, 277-288, 2002.

[Law03]
J. H. Lawton, Opportunism in Planning Agents, *The Seventh World Multiconference on*

Systemics, Cybernetics and Informatics, July, 2003.

[LD03a]
C. Domshlak and J. H. Lawton, On Planning for Multi-Agent Opportunistic Execution, *IJCAI-03 Workshop on Issues in Designing Physical Agents for Dynamic Real-Time Environments*, 2003.

[LD03a]
J. H. Lawton and C. Domshlak, Towards Multi-Agent Opportunism with Planning Agents, *AAMAS-03 Workshop on Autonomy, Delegation, and Control*, 2003.

[LD04a]
J.H. Lawton and C. Domshlak, On the Role of Knowledge in Multi-Agent Opportunism, *Third International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2004.

[MIDM03]
H. Kautz, AFRL/IISI Workshop on Mixed Initiative Decision Making: Final report, 2003, *http://www.cis.cornell.edu/iisi/MIDM-workshop/final%20report/midm-report.htm.*

[Pat91]
A. Patalano, C. Seifert, and K. Hammond, Predictive Encodings: Planning for Opportunities, Proceedings of the 15th Conference of the Cognitive Science Society, 800--805, 1993.

[Wool99]
M. Wooldridge, Intelligent Agents, *Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence*, ed. Gerhard Weiss, The MIT Press, Cambridge, MA, 1999

4. List of IISI Research Projects

Examples of collaborative research projects between IISI and outside research institutions:

CORE – Computational Principles for Optimization of Resources and Execution Time, a joint project involving IISI, Microsoft Research, and University of Washington.

CATS – Combinatorial Auctions Test Suite, a joint project involving IISI, University of Washington, and Stanford University.

MUSA – Multi-Valued Satisfiability, a joint project involving IISI and University of Barcelona.

ACTIVITY COMPASS - a joint project involving IISI and the University of Washington.

Examples of Cornell projects sponsored by IISI:

Active Memory Clusters

Balanced Latin Squares for Experimental Design

Controlling Computational Cost: Structure and Phase Transition Phenomena

Data-mining

Discovering Hidden Structure from Heterogeneous Information Sources

Dual Miner

Distributed CSP

Meta Clustering

RoboCup and RoboFlag

Thread Migration on Clusters of Workstations

5. Examples of conferences and meetings organized and sponsored by IISI

7th International Conference on Theory and Applications of Satisfiability Testing (SAT 2004)

Human Language Technology Conference/ North American Chapter of the Association for Computational Linguistics Annual Meeting (HLT/NAACL 2004)

6th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CP-AI-OR 2004)

AFRL/IISI Workshop on Mixed Initiative Decision Making (2003)

9th International Conference on Principles and Practice of Constraint Programming (CP 2003)

18th International Joint Conference on Artificial Intelligence (IJCAI 2003)

IISI Workshop on Strategic Research Directions in AI (2003)

6th International Conference on Theory and Applications of Satisfiability Testing (SAT 2003)

5th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CP-AI-OR 2003)

8th International Conference on Principles and Practice of Constraint Programming (CP 2002)

Symposium On Using Uncertainty within Computation, Fall Symposium series, under the auspices of the American Association of Artificial Intelligence, North Falmouth, Cape Cod (2001)

Meeting of the North American Chapter of the Association for Computational Linguistics (2001)

School on Optimization at the 4th International Workshop on the Integration of AI and OR techniques in Constraint Programming for Combinatorial Optimization Problems, Nice, France (2002)

5th International Symposium on the Theory and Applications of Satisfiability Testing, Cincinnati, OH (2002)

School on Statistical Physics, Probability Theory and Computational Complexity, at The International Centre for Theoretical Physics, Trieste, Italy (2002)

Conference on Typical-Case Complexity, Randomness, and Analysis of Search Algorithms, The International Centre for Theoretical Physics, Triest, Italy (2002)

6. Personnel Supported

Carlos Ansotegui (Computer Science)
Krishna B. Athreya (Operations Research)
Claire Cardie (Computer Science)
Rich Caruana (Computer Science)
Raffaello D'Andrea (Mechanical & Aerospace Engineering)
Carmel Domshlak (Computer Science)
Johannes Gehrke (Computer Science)
Carla Gomes (Computer Science)
Joseph Halpern (Computer Science)
Juris Hartmanis (Computer Science)
John Hopcroft (Computer Science)
Thorsten Joachims (Computer Science)
Jon Kleinberg (Computer Science)
Lillian Lee (Computer Science)
David Schwartz (Computer Science)

Meinolf Sellmann (Computer Science)
Bart Selman (Computer Science)
Phoebe Sengers (Information Science)
David Shmoys (Operations Research)
Chris Shoemaker (Civil Engineering)
Steve Strogatz (Theoretical and Applied Mechanics)
Stephen Wicker (Electrical and Computer Engineering)

In addition, several students were also supported by IISI.

6. Interactions/Transitions

A key component of IISI's mission is to foster research collaborations with AFRL/IF and the research community in general. The past three years have been very active in research collaborations.

a. AFRL/IF Researchers

IISI had collaborations and research interactions with the following AFRL/IF researchers:

Joe Carozzoni
Jerry Dussault
Nate Gemelli
James Lawton
Mark Linderman
Bob Paragi
Louis Pochet
Nancy Roberts
Peter Rocci
Justin Sorice
John Spina
Matt Thomas
Robert Wright

Nancy Roberts obtained her MSc's degree. For her thesis, she worked with Carla Gomes and Mike Pitarelli. James Lawton is currently working towards his PhD. Carla Gomes and Carmel Domshlak from IISI are advising him.

b. Other Researchers

IISI has hosted several visitors from different research institutions:

Dimitris Achlioptas (Microsoft Research)

Shai Ben-David (Technion, Israel)
Ronen Brafman (Ben Gurion Un.)
Pedro Domingos (U. Washington)
Cesar Fernandez (University of Lleida)
Russ Grenier (U. Alberta)
Eric Horvitz (Microsoft Research)
Kalev Kask (UC Irvine)
Henry Kautz (U. Washington)
Scott Kirkpatrick (Hebrew Un.)
Yann LeCun (NYU)
Kevin Leyton-Brown (U. British Columbia)
Michael Littman (Rutger Un.))
Felip Manya (University of Lleida)
David McAllester (ATT Labs)
Andrew McCallum (U. Massachusetts)
Fernando Pereira (U. Pennsylvania)
Jean-Charles Regin (ILOG/CPLEX)
Yoav Shoam (Stanford University)
Manuela Veloso (CMU)
Toby Walsh (York University, UK)
Walker White (U. Dallas)
Wayne Zhang (Un. Washington, St. Louis)

c. Presentations by IISI Members

"A priori generalization bounds for SVM's - can Sparsity save the day?", NIPS workshop on Kernel Methods, NIPS 2001, Vancouver Canada, December 2001 (Shai Ben-David)

"A theoretical framework for learning from a pool of disparate data sources", KDD 2002 conference, Edmonton, Alberta, Canada, July 2002 (Shai Ben-David)

"Active Memory Clusters: Efficient Multiprocessing on Commodity Clusters", Seminar, School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL, April 2002; International Symposium on High-Performance Computing (ISHPC), Kansai Science City, Japan, May 2002 (Mark Heinrich)

"Active Memory Systems: A Unified Uniprocessor and Multiprocessor Approach", Seminar, IBM Austin Research Labs, Austin, TX, December 2001 (Mark Heinrich, Evan Speight)

"Active Memory Systems Research", Seminar, IBM Research, Yorktown, NY, September 2001 (Mark Heinrich, Evan Speight)

"Bayesian predictive model of an interactive environment", Presentation to AFRL/IFT Division, May 2002 (Nancy Roberts)

“Bias Correction for Decision Tree Construction”, Computer Science Colloquium, Department of Computer and Information Science, Polytechnic University, Brooklyn, New York, December 2001; Invited Talk at AT&T Research as part of the Dimacs Special Year on Data Mining. Florham Part, New Jersey, January 2002 (Johannes Gehrke)

"Cache Coherence Protocol Design for Active Memory Systems", International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA), Las Vegas, NV, June 2002 (Mark Heinrich)

“Combining Sample Selection and Error-Driven Pruning for Machine Learning of Coreference Rules”, Conference on Empirical Methods in NLP, 2002 (Vincent Ng)

“Communication and computation in DCSP algorithms”, C. Fernández, R. Béjar, B. Krishnamachari and C. Gomes, 8th International Conference on Principles and Practice of Constraint Programming, CP'2002, Ithaca, NY, September 2002 (Cesar Fernandez)

“Control of Complex Systems”, The Center for Bits and Atoms, Massachusetts Institute of Technology, Boston, MA, October 2002 (Raffaello D’Andrea)

“Cooperative Vehicle Control”, University of Florida, Graduate Engineering Research Center, Research Institute for Autonomous Precision Guided Systems, Shalimar, FL, May 2002; GRASP Laboratory, University of Pennsylvania, Philadelphia, PA, May 2002 (Raffaello D’Andrea)

“Data Management for Sensor Networks”, Invited talk at the Air Force Research Laboratory, Rome, NY, June 2002 (Johannes Gehrke)

"Delphi: Prediction-Based Page Prefetching to Improve the Performance of Shared Virtual Memory Systems", International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, June, 2002 (Evan Speight)

“Distributed Mining and Monitoring”, Knowledge Discovery and Dissemination (KD-D) Workshop, Leesburg, Virginia, December 2001; First Knowledge Discovery and Dissemination Workshop, Leesburg, Virginia, December 2001; Knowledge Discovery and Dissemination Discussion Meeting, Reston, Virginia, January 2002; Invited talk at the Dimacs Workshop on Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense, Princeton, NJ, June 2002; Invited talk at the University of Washington-Seattle. Seattle, WA, July 2002 (Johannes Gehrke)

"Experiences in Designing Experiences", CHI 2002 Workshop on Funology, Minneapolis, MN, April 2002 (Phoebe Sengers)

"Evaluating Search Engines using Clickthrough Data", SIGIR Conference, Workshop on Mathematical Models, Tampere, Finland, August 2002 (Thorsten Joachims)

“Exploiting Structure and Randomization in Combinatorial Search”, School on Optimization, CP-AIOR 2002, Le Croisic, France, March 2002 (Carla Gomes)

“Extending the Reach of SAT with Many-valued Logics”, R. Béjar, A. Cabiscol, C. Fernández, C. P. Gomes and F, Manyà, Workshop on Theory and Applications of Satisfiability, IEEE Symposium on Logic in Computer Science, Boston (MA), USA, 2001 (Cesar Fernandez)

“GADT: A Probability Space ADT For Representing and Querying the Physical World”, Paper presentation at the 18th International Conference on Data Engineering (ICDE 2002), San Jose, California, February 2002 (Johannes Gehrke)

"Generalized Graph Cuts for Transduction", Fraunhofer Research Center, Institute FIRST, Berlin, Germany, June 2002 (Thorsten Joachims)

“Generating hard satisfiable scheduling instances”, J. Argelich, R. Béjar, A. Cabiscol, C. Fernández, F. Manyà and C. Gomes, 6th European Conference on Planning, ECP'2001, Toledo, Spain, 2001 (Cesar Fernandez)

“Heavy-tailed phenomena in combinatorial search”, Santa Fe Institute, Sept. 2001 (Carla Gomes)

“Heuristic Algorithms: Theory and Practice”, Summer School on Statistical Physics, Probability Theory, and Computational Complexity, The International Centre for Theoretical Physics, Trieste, Italy, Sep. 2002 (Carla Gomes)

“Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution”, Nineteenth International Conference on Computational Linguistics (COLING-02), 2002 (Vincent Ng)

“Improving Machine Learning Approaches to Coreference Resolution”, Fortieth Anniversary Meeting of the Association for Computational Linguistics (ACL-02), 2002 (Vincent Ng)

“Improving Machine Learning Approaches to Noun Phrase Coreference Resolution”, University of Texas at Austin, February 2002 (Claire Cardie)

“Information Routing in Wireless ad-hoc Networks”, The Information Awareness and Understanding Branch (AFRL), May 2002; Terry Lyons of Asian Office of Aerospace Research and Development (AOARD), May 2002 (John Spina)

“Knowledge-Lean Approaches to Statistical Natural Language Processing”, Cornell Cognitive Studies Symposium on Statistical Learning Across Cognition, April 2002

"Learning from a pool of disparate data sources", Invited talk at the Annual NeuroCOLT Workshop at Cumberland Lodge, London, UK, April 2002 (Shai Ben-David)

"Learning Ranking Functions from Preference Data", University of Eindhoven, EURANDOM Center for Statistics, Netherlands, June 2002 (Thorsten Joachims)

"Learning Text Classifiers with Support Vector Machines", University College Dublin, Computer Science Colloquium Series, Ireland, January 2002; Universitaet Stuttgart, Institut fuer Maschinelle Sprachverarbeitung, Germany, January 2002 (Thorsten Joachims)

"Leveraging Cache Coherence in Active Memory Systems", 16h International Conference on Supercomputing (ICS), NYC, June 2002. (Daehyun Kim: Winner, Best Student Presentation)

"Optimizing Search Engines using Clickthrough Data", Daimler-Chrysler Research, Ulm, Germany, May 2002; Fraunhofer Research Center, Institute FIRST, Berlin, Germany, June 2002; Conference on Knowledge Discovery in Databases (KDD), Edmonton, Canada, July 2002 (Thorsten Joachims)

"PermaNT: Persistent Shared Memory for Windows NT/2000 Clusters", International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, June, 2002 (Evan Speight)

“Phase Transitions and Structure in Combinatorial Problems”, American Association for Artificial Intelligence (AAAI), Edmonton, Canada, July, 2002 (Carla Gomes)

“Probabilistic Target Tracking with a Network of Distributed Sensor Agents”, IISI Meeting, March 2002 (Matthew Thomas)

“Querying the Physical World”, IBM Almaden Research Center, San Jose, California, February 2002; Lockheed Martin, Owego, New York, March 2002; DARPA Information Exploitation Office, Arlington, Virginia, May 2002 (Johannes Gehrke)

“Randomization and Rational Decision Making in Optimization”, Uncertainty in Artificial Intelligence (UAI), Edmonton, Alberta, Canada, 2002 (Carla Gomes)

“Randomization, Structure, and Complexity in Combinatorial Optimization”, Institute for Pure and Applied Mathematics, UCLA, April 2002 (Carla Gomes)

“Research Questions in Data Management for Sensor Networks”, NSF Workshop on Context-Aware Mobile and Pervasive Data Management, Providence, Rhode Island, January 2002 (Johannes Gehrke)

“Some New Thoughts On Old Questions For Even Older Data Mining Problems”, Stanford University, Palo Alto, California, March 2002 (Johannes Gehrke)

"Stupid, but Lovable: How Formal Structures Create Human Meaning", Swedish Institute of Computer Science, June 2002 (Phoebe Sengers)

“Tasking Sensor Networks”, DARPA Sensor Information Technology PI Meeting, Santa Fe, New Mexico, January 2002 (Johannes Gehrke)

“The Integration of Constraint Programming and Mathematical Programming Methods”, Institute for Pure and Applied Mathematics, UCLA, June 2002 (Carla Gomes)

“The Iterative Residual Rescaling Algorithm: An Analysis and Generalization of Latent Semantic Indexing”, University of California, San Diego, January, 2002

"Using Bayesian Networks and Decision Theory to Model Security", MS in Computer Science Presentation to SUNY faculty, December 2001 (Nancy Roberts)

“Vision and Directions for the Intelligent Information Systems Institute”, AFRL/IF, Scientific Advisory Board, Nov. 2001 (Carla Gomes)

7. Publications by IISI Members

Athreya, K.B. (2004): Stationary Measures for some Markov chain models in ecology and economics., Economic Theory, 23, 107-122.

Athreya, K.B. (2004): Markov chains generated by iid random perturbations of deterministic interval maps, Tech report, ORIE, Cornell. (to appear in Annals of Probability).

Athreya, K.B. (2004): Markov chains generated by iid random maps on Polish spaces, Tech report, ORIE, Cornell, to appear in Kolmogorov's birth centenary volume, University of Paris.

Athreya, K.B. (2004): Non negative Markov chains with applications. Proceedings of the IMA conference on Probability and PDE, University of Minnesota and Springer Verlag.

Athreya, K.B., Krishnan, T. and Delampady, M. (2003-2004) (four papers) Markov Chain Monte Carlo-I, II, III, IV. in Resonance, Journal of Science Education, Indian Academy of Sciences, Bangalore, India.

Athreya, K B. and Gomes, C.P. (2004): Heavy tailed distributions with bounded support: Applications to combinatorial search, working paper, IISI, Cornell University.

Athreya, K.B. (2004): From olympiad problem to the maximum principle, Resonance, (2004).

Athreya, K.B. and Schuh, H.J. (2003): Random logistic maps-II, the critical case, Journal of theoretical Probability, volt 16, #4, 813-830.

Athreya, K.B. and Stenflo, O. (2003): Perfect sampling for Doeblin chains., to appear in Sankhya.

Athreya, K.B. and Majumdar, M. (2003): Estimating the stationary distribution of a Markov chain., Economic Theory, 21,729-742.

Athreya, K.B., Hitchcock, J., Lutz, J. and Mayordamo, E. (2003): Effective strong dimension, algorithmic information and computational complexity, Tech report, ORIE, Cornell, to appear in SIAM journal on computing.

Athreya, K.B. and Dai, J. (2002): On the nonuniqueness of invariant probability measure for random logistic maps., Annals of Probability, 30,1142-1147

Athreya, K.B. (2002): Harris irreducibility of iterates of IID random maps on R+.Tech report, ORIE, Cornell.

Ayres, Jay; Gehrke, J. E.; Yiu, Tomi; and Flannick, Jason. "Sequential PAttern Mining Using Bitmaps". *In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, July 2002.

Barzilay, Regina and Lillian Lee. Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization. *Proceedings of HLT/NAACL 2004* (to appear).

Barzilay, Regina and Lillian Lee. Learning to Paraphrase: An Unsupervised Approach using Multiple-Sequence Alignment. *Proceedings of HLT/NAACL 2003, pp. 16--23.*

Bejar R., Domshlak C., Fernandez C., Gomes C., Krishnamachari B., Selman B., Valls M., Sensor networks and distributed CSP: Communication, Computation and Complexity. *Artificial Intelligence Journal.* Accepted for publication.

Bejar, Ramon; Cabiscol, Alba; Fernandez, Cesar; Manya, Felip; and Gomes, Carla. "Capturing the Structure of Satisfiability". *Proceedings of 7th Intl. Conference on the Principles and Practice of Constraint Programming (CP-2001)*, 2001, 137–153.

Bejar, Ramon; Cabiscol, Alba; Fernandez, Cesar; Manya, Felip; and Gomes, Carla. "Extending the Reach of SAT with Many-Valued Logics". *Electronic Notes in Discrete Mathematics*, Vol. 9, Elsevier Science Publ., 2001.

Bejar, Ramon; Cabiscol, Alba; Fernandez, Cesar; Manya, Felip; and Gomes, Carla. "Regular-SAT: A Many-Valued Approach for Solving Combinatorial Problems". *Discrete Applied Mathematics*, Elsevier. Forthcoming.

Ben-David, Shai. "A Priori Generalization Bounds for Kernel Based Learning". Invited to a special issue of the *Journal of Machine Learning Research.*

Ben-David, Shai; Gehrke, Johannes; and Schuller, Reba. "A Theoretical Framework for Learning from Disparate Data Sources". *Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Ben-David, Shai; and Schuller, Reba. "Exploiting Task Relatedness for Multiple Task Learning". Submitted to the upcoming *NIPS'02 Conference.*

Bessiere, C.; Fernandez C.; Gomes C.; and Valls, M., Pareto-like Distributions in Random Binary CSP. *Frontiers in Artificial Intelligence and Applications - Artificial Intelligence Research and Development*, IOS Press, Vol 100 ISSN 0922-6389, 451-461, 2003.

Boutilier, C., R. Brafman, C. Domshlak, D. Poole, and H. Hoos, CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements. Journal of Artificial Intelligence Research, Vol. 21, 135-191, 2004.

Brafman, R. and C. Domshlak, Structure and Complexity of Planning with Unary Operators. Journal of Artificial Intelligence Research, Vol. 18, 315-349, 2003.

Bucila, Cristian, J.E. Gehrke, Daniel Kifer, Walker White. DualMiner: A Dual-Pruning Algorithm for Itemsets with Constraints. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2002).* Edmonton, Alberta, Canada, July 2002. Invited to *Data Mining and Knowledge Discovery,* Special Issue on "Best Papers of KDD 2002".

Bucila, Cristian, J. E. Gehrke, Daniel Kifer, and Walker White. DualMiner: A Dual-Pruning Algorithm for Itemsets with Constraints. *Data Mining and Knowledge Discovery*, Vol. 7, No. 3, July 2003, pages 241-272. Invited paper for the Special Issue on "Selected Papers from the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—Part I."

Cardie, Claire, Janyce Wiebe, Theresa Wilson, and Diane Litman. Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. *2003 AAAI Spring Symposium on New Directions in Question Answering*, 20–27, AAAI Press, 2003.

Caruana, Rich and de Sa, Virginia R., "Benefiting from the Variables that Variable Selection Discards," Journal of Machine Learning Research (JMLR), Vol. 3, March 2003, pp.1245-1264.

Caruana, Rich, Niculescu, Stefan, Rao, Bharat, and Simms, Cynthia, "Evaluating the C-section Rate of Different Physician Practices: Using Machine Learning to Model

Standard Practice" to appear at the American Medical Informatics Conference (AMIA), November 2003.

Charikar, M., S. Guha, E. Tardos, and D.B. Shmoys. "A constant-factor approximation algorithm for the k-median problem". Journal Computer System Sciences 65, 2002, 129-149.

Chaudhuri, M.; Kim, D.; and Heinrich, M. "Cache Coherence Protocol Design for Active Memory Systems". *In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications*, June 2002, 83-89.

Chen, Hubie; Gomes, Carla; and Selman, Bart. "Formal Models of Heavy-tailed Behavior in Combinatorial Search". *Proceedings of 7th Intl. Conference on the Principles and Practice of Constraint Programming (CP-2001)*, 2001, 408–422.

Chen, J., X. Zhang, T. Berger and S. B. Wicker, "The Sum-Rate Distortion Function and Optimal Rate Allocation for the Quadratic Gaussian CEO Problem" to appear in the IEEE Journal on Selected Areas in Communications: Special Issue on Sensor Networks.

Chen, J., X. Zhang, T. Berger and S. B. Wicker "Rate Allocation in Distributed Sensor Network," Proceedings of the Allerton Conference 2003.

Chen, Yurong and Stephen Wicker, "On Selection of Optimal Transmission Power for Ad hoc Networks," ACM/Baltzer Wireless Networks, 2002.

Chen, Yurong and Stephen B. Wicker, On Selection of Optimal Transmission Power for Ad hoc Networks, Proceedings of the Hawaii International Conference on System Science (HICSS-36), Big Island, Hawaii, Jan. 6-9, 2003.

D'Andrea, R. and G. E. Dullerud. Distributed Control Design for Spatially Interconnected Systems. IEEE Transactions on Automatic Control, 48(9):1478--1495, 2003.

D'Andrea, R. Temporal Discretization of Spatially Interconnected Systems. Automatica. Submitted for publication.

D'Andrea, R. and R. S. H. Istepanian. Design of Full State Feedback Finite Precision Controllers. International Journal of Robust and Nonlinear Control, 12:537--553, 2002.

Dantsin, E., A. Goerdt, E. Hirsch, R. Kannan, J. Kleinberg, C. Papadimitriou, P. Raghavan, U. Schoning, A deterministic algorithm for satisfiability based on local search. Theoretical Computer Science, 289(2002).

Domshlak, C. and S. E. Shimony, Complexity of Probabilistic Reasoning in Directed-Path Singly Connected Bayes. *Networks* Artificial Intelligence, Vol. 151(1-2), 213-225, 2004.

Domshlak, C. and J. Lawton, The Importance of Knowledge in Multi-Agent Opportunism. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems, 2004.

Domshlak, C., F. Rossi, B. Venable, and T. Walsh, Reasoning about Soft Constraints and Conditional Preferences. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-03), 215-220, 2003.

Domshlak, C. and S. E. Shimony, Efficient Probabilistic Reasoning in Bayes Nets with Mutual Exclusion and Context Specific Independence, In Proceedings of FLAIRS-03, Special Track on Uncertainty in AI.

Earl, M.G. and S.H. Strogatz. Synchronization in oscillator networks with delayed coupling: A stability criterion. Physical Review E 67, 036204 (2003).

Earl, M. G.; and D'Andrea, R. "A study in cooperative control: the Robo Flag drill". *Proceedings of American Control Conference 02 (ACC 02)*, 2002, 1811-1812.

Earl, M. G.; and D'Andrea, R. "Modeling and control of a multi-vehicle system using mixed integer linear programming". Proceedings of Conference on Decision and Control (CDC 02), To Appear, 2002.

Felzenszwalb, P., D. Huttenlocher, J. Kleinberg. Fast Algorithms for Large-State-Space HMMs with Applications to Web Usage Analysis. Advances in Neural Information Processing Systems (NIPS) 16, 2003.

Fernandez, Cesar; Bejar, Ramon; Krishnamachari, Bhaskar; and Gomes, Carla. "Communication and Computation in DisCSP Algorithms". *Proceedings of 8th Intl. Conference on the Principles and Practice of Constraint Programming (CP-2002)*, 2002, 664–679.

Ginsparg, P., J. Gehrke, J. Kleinberg. Overview of the 2003 KDD Cup. SIGKDD Explorations, 2004.

Girvan, M., D.S. Callaway, M.E.J. Newman, and S.H. Strogatz. A simple model of epidemics with pathogen mutation. Physical Review E 65, 031915 (2002).

Goldberg, D., S. McCouch, J. Kleinberg. Constructing comparative maps with unresolved marker order. Proc. Pacific Symposium on Biocomputing, 2002.

Goldenberg, A., Shmueli, G., Caruana, R., Fienberg, S., "Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-counter Medication Sales," *Proceedings of the National Academy of Sciences*, 99, 5237-5240, 2002.

Goldsmith, Andrea and Stephen B. Wicker, "Design Challenges For Energy-Constrained Ad Hoc Wireless Networks," IEEE Wireless Communications Magazine, August, 2002.

Gomes, Carla, Meinolf Sellmann, Cindy van Es, and Harold van Es, The Challenge of Generating Spatially Balanced Scientific Experiment Designs, Accepted at CP-AI-OR, 2004.

Gomes, Carla, Xi Xie, Stephen B. Wicker, and Bart Selman, "Heavy Tails, Phase Transitions, and the Nature of Cutoff," in Codes, Graphs, and Systems, Boston: Kluwer, 2002.

Gomes, Carla. "On the Intersection of Artificial Intelligence and Operations Research". *Journal of Knowledge Engineering Review*, Cambridge Press, Vol. 16 1, 2001, 1–6.

Gomes, Carla; and Kautz, Henry. "Completing Latin Squares: A Challenge Problem in Combinatorial Optimization". *Discrete Applied Mathematics*, Elsevier. Forthcoming.

Gomes, Carla; Regis, Rommel; and Shmoys, David. "An improved approximation algorithm for the partial latin square extension problem". *Proceeding of the ACM-SIAM Symposium on Discrete Algorithms (SODA-2003)*, San Francisco, CA, 2003. Forthcoming.

Gomes, Carla; and Selman, Bart. "Algorithm Portfolios". *Artificial Intelligence Journal*, Vol. 126, 2001, 43–62.

Gomes, Carla; and Selman, Bart. "Satisfied with Physics". *Science*, Vol. 297, Aug. 2, 2002, 784–785. (Perspective article.)

Gomes, Carla, Shmoys, David. Approximations and Randomization to Boost CSP Techniques. *Annals of Operations Research.* Accepted for publication.

Gomes, Carla; Williams, R. Approximation Algorithms. In *Introduction to Optimization, Decision Support and Search Methodologies*, Burke and Kendall (Eds.), Kluwer, Forthcoming.

Gomes, Carla. Solving Hard Computational Problems Using Complete Randomized Search Methods. *AI Magazine*. Invited survey. Forthcoming.

Gomes, Carla; Regis, Rommel; Shmoys, David. An Improved Approximation Algorithm for the Partial Latin Square Extension Problem. *Proceeding of the ACM-SIAM Symposium on Discrete Algorithms (SODA-2003)*, Baltimore, 2003.

Gomes, Carla. Complete Randomized Backtrack Search. In *Constraint and Integer Programming: Toward a Unified Methodology*, Milano, M., (ed.), Kluwer, 2003, 233–283.

Gomes, Carla and Shmoys, David. The Promise of LP to Boost CSP Techniques for Combinatorial Problems. In *Proceedings of the 4th International Symposium on Integration of AI and OR Techniques in Constraint Programming for Combinatorial*

Optimization Problems (CP-AI-OR'02), 291-305, Le Croisic, France, 291-305, March 2002.

Gomes, Carla; and Selman Bart. Hill Climbing Search. In *Nature Encyclopedia of Cognition*, Nature Publ., 2002.

Gomes, Carla. Hybrid Compute Intensive Approaches for Combinatorial Optimization. Technical Report, RL-TR-02-65, AFRL, Information Directorate, 2002.

Halpern, Joe; Chu, F.; and Gehrke, J. "Least expected cost query optimization: What can we expect?". *Proceedings of the 21st ACM Symposium on Principles of Database Systems*, 2002, 293-302.

Halpern, Joe; Grunwald, P. "Updating probabilities". *Proceedings of the Eighteenth Conference on Uncertainty in AI*, 2002, 187-196.

Halpern, Joe; Haas, Z.; and Li, L. "Gossip-based ad hoc routing". *Proceedings of Infocom*, 2002, 1707-1716.

Halpern, Joe; and O'Neill, K. "Secrecy in multi-agent systems". *Proceedings of the 15th IEEE Computer Security Foundations Workshop*, 2002, 32-46.

Halpern, Joe; Pucella, R. "A logic for reasoning about upper probabilities". *Journal of AI Research* 17, 2002, 57-81.

Halpern, Joe; Pucella, R. "Reasoning about expectation". *Proceedings of the Eighteenth Conference on Uncertainty in AI*, 2002, 207-215.

Halpern, Joseph, A logical reconstruction of SPKI, to appear, *Journal of Computer Security* (with R. van der Meyden).

Halpern, Joseph, A cone-based distributed topology-control algorithm for wireless multi-hop networks, to appear, *IEEE/ACM Transactions on Networks*, (with L. Li, M. Bahl, Y. Wang, R. Wattenhofer).

Halpern, Joseph, Complete axiomatizations for reasoning about knowledge and time, to appear, *SIAM Journal on Computing* (with R. van der Meyden and M. Vardi).

Halpern, Joseph, *Reasoning About Uncertainty*, MIT Press, 2003.

Halpern, Joseph, A computer scientist looks at game theory, *Games and Economic Behavior* 45:1, 2003, pp. 114-131.

Halpern, Joseph, On the relationship between strand spaces and multi-agent systems, *ACM Transactions on Information and System Security* **6**:1, 2003, pp. 43-70 (with R. Pucella).

Halpern, Joseph, Updating probabilities, *Journal of AI Research* 19, 2003 (with P. Grunwald).

Halpern, Joseph, Characterizing the common prior assumption, *Journal of Economic Theory*, 106:2, 2002, pp. 316--355.

Halpern, Joseph, Using first-order logic to reason about policies, to appear, *Proceedings of the 16th IEEE Computer Security Foundations Workshop*, 2003, pp. 187-201 (with V. Weissman).

Halpern, Joseph, Anonymity and information hiding in multiagent systems, *Proceedings of the 16th IEEE Computer Security Foundations Workshop*, 2003, pp. 75-88 (with K. O'Neill).

Halpern, Joseph, Great Expectations. Part I: On the Customizability of Generalized Expected Utility, *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003)*, 2003, pp. 291-296 (with F. Chu).

Halpern, Joseph, Great Expectations. Part II: Generalized Expected utility as a universal decision rule, *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003)*, 2003, pp. 297-302 (with F. Chu).

Halpern, Joseph, Responsibility and blame: A structural-model approach, *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003)*, 2003, pp. 147-153 (with H. Chockler).

Halpern, Joseph, Probabilistic algorithmic knowledge, *Proceedings of the Ninth Conference on Theoretical Aspects of Rationality and Knowledge*, 2003, pp. 118-130 (with R. Pucella).

Halpern, Joseph, A Logic for Reasoning about Evidence, *Proceedings of the Nineteenth Conference on Uncertainty in AI*, 2003, pp. 297-304 (with R. Pucella).

Halpern, Joseph, Editorial: *JACM*'s 50th Anniversary, *Journal of the ACM* 50:1, 2003, pp.3 -7.

Halpern, Joseph, Modeling adversaries in a logic for security protocol analysis, *Proceedings: Formal Aspects of Security*, 2002 (with R. Pucella).

Halpern, Joseph, Update: Time to Publication Statistics, *Journal of the ACM* 49:6, 2002, p. 715.

Heinrich, M.; Speight, E.; and Chaudhuri, M. "Active Memory Clusters: Efficient Multiprocessing on Commodity Clusters". *In Proceedings of the 4th International Symposium on High-Performance Computing, Lecture Notes in Computer Science*, Springer-Verlag, Vol. 2327, May 2002, 78-92.

Hopcroft, John, Brian Kulis, Omar Khan, and Bart Selman. Tracking evolving communities in large linked networks. *Proc. Natl. Acad. of Sci.* (PNAS), Feb., 2004.

Hopcroft, John, Brian Kulis, Omar Khan, and Bart Selman. Natural communities in large linked networks. *Proc. KDD,* August, 2003.

Hueffmeier, Ewald, Janak Sodha, and Stephen B. Wicker, "On the Termination of the BCJR Algorithm," Proceedings of the Third International Symposium on Communication Systems, Networks And Digital Signal Processing, Staffordshire, England, 15-17 July 2002.

Joachims, T., Transductive Learning via Spectral Graph Partitioning, Proceedings of the International Conference on Machine Learning (ICML), Morgan Kaufmann, 2003

Joachims, T., Unbiased Evaluation of Retrieval Quality using Clickthrough Data, in R. Nakhaeizadeh, Text Mining, Springer, 2003.

Joachims, T., Optimizing Search Engines Using Clickthrough Data, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2002

Joachims, T. "Learning to Classify Text using Support Vector Machines, Kluwer, 2002.

Joachims, T. "Optimizing Search Engines Using Clickthrough Data". *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2002.

Joachims, T. "Unbiased Evaluation of Retrieval Quality using Clickthrough Data", in *R. Nakhaeizadeh, Text Mining*, Springer, 2002.

Joachims, T. "A Statistical Learning Model of Text Classification with Support Vector Machines". *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR)*, ACM, 2001.

Jun, M.; Chaudhry, A.; and D'Andrea, R. "The navigation of autonomous vehicles in uncertain dynamic environments: A case study". *Proceedings of Conference on Decision and Control (CDC 02)*, 2002.

Jun, M.; and D'Andrea, R. "Path planning for unmanned aerial vehicles in uncertain and adversarial environments". ed. S. Butenko and R. Murphey and P. Pardalos. *Cooperative Control: Models, Applications and Algorithms*, Kluwer, 2002, 95-111.

Kalmár-Nagy, T., T., R. D'Andrea, and P. Ganguly. Near-Optimal Dynamic Trajectory Generation and Control of an Omnidirectional Vehicle. Robotics and Autonomous Systems, 46: 47--64.

Kautz, Henry; Horvitz, Eric; Ruan, Yongshao; Gomes, Carla; and Selman, Bart. “Dynamic Restart Policies”. *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, Edmonton, Alberta, Canada, 2002, 674–682.

Ke, J.; and Speight, E. “Tern: Migrating Threads in an MPI Runtime Environment”. Submitted to the *International Parallel and Distributed Processing Symposium*, September 2002.

Ke, J.; and Speight, E. “The Design and Performance of the Tern Thread Migration Systems”. Submitted to the *Journal of Parallel and Distributed Computing*, 2002.

Kifer, Daniel, J. E. Gehrke, Cristian Bucila, and Walker White. How to Quickly Find a Witness. In *Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS 2003).* San Diego, CA, June 2003.

Kim, D.; Chaudhuri, M.; and Heinrich, M. “Active Memory Techniques for ccNUMA Multiprocessors". Submitted to the *International Symposium on High-Performance Computer Architecture*, July 2002.

Kim, D.; Chaudhuri, M.; and Heinrich, M. “Leveraging Cache Coherence in Active Memory Systems”. *In Proceedings of the 16th ACM International Conference on Supercomputing*, June 2002, 78-92.

Kleinberg, J., M. Sandler. Using Mixture Models for Collaborative Filtering. Proc. 36th ACM Symposium on Theory of Computing, 2004.

Kleinberg, J., M. Sandler, A. Slivkins. Network Failure Detection and Graph Connectivity. Proc. 15th ACM-SIAM Symposium on Discrete Algorithms, 2004.

Kleinberg, J., M. Sandler. Convergent Algorithms for Collaborative Filtering. Proc. 4th ACM Conference on Electronic Commerce, 2003.

Kleinberg, J. Bursty and Hierarchical Structure in Streams. Proc. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2002.

Kleinberg, J. An Impossibility Theorem for Clustering. Advances in Neural Information Processing Systems (NIPS) 15, 2002.

Krishnamachari, B. and S. B. Wicker, "Base Station Location Optimization in Cellular Wireless Networks using Heuristic Search Algorithms," book chapter in Soft Computing in Communications, Ed. L. Wang, Springer-Verlag, 2004.

Krishnamachari, Bhaskar, Rung-Hung Gau, Stephen B. Wicker, and Zygmunt J. Haas, "Optimal Sequential Paging in Cellular Networks," ACM/Baltzer Wireless Networks, Vol 10, No. 2, March 2004.

Krishnamachari, Bhaskar, Stephen Wicker, Ramon Bejar and Cesar Fernandez, "On the Complexity of Distributed Self-Configuration in Wireless Networks," in Kluwer Journal on Telecommunication Systems, Special Issue on Wireless Networks and Mobile Computing, Eds. I. Stojmenovic and S. Olariu, Vol. 22, No. 1, January/April 2003.

Krishnamachari, Bhaskar, Stephen B. Wicker, Ramon Bejar, and Marc Pearlman, "Critical Density Thresholds in Distributed Wireless Networks," Advances in Coding and Information Theory, eds. H. Bhargava, H.V. Poor and V. Tarokh, Kluwer Publishers, 2002.

Krishnamachari, Bhaskar, Deborah Estrin, Stephen Wicker, ``The Impact of Data Aggregation in Wireless Sensor Networks," International Workshop on Distributed Event-Based Systems, (DEBS '02), held in conjunction with IEEE ICDCS, Vienna, Austria, July 2002.

Krishnamachari, Bhaskar, Ramon Bejar, and Stephen B. Wicker, "Distributed Problem Solving and the Boundaries of Self-Configuration in Multi-hop Wireless Networks" Hawaii International Conference on System Sciences (HICSS-35), January 2002.

Kubota Ando, Rie and Lillian Lee. Mostly-Unsupervised Statistical Segmentation of Japanese Kanji Sequences (pre-publication version). *Natural Language Engineering* 9(2), pp. 127--149, 2003.

Kurland, Oren and Lillian Lee. Corpus structure, language models, and ad-hoc information retrieval. *Proceedings of SIGIR 2004* (to appear).

Langford, John, and Caruana, Rich, "(Not)Bounding the True Error," *Neural and Information Processing Systems*, Vol. 14 (Proceedings of NIPS*2001), MIT Press, 2002.

Lee, Lillian. "I'm sorry Dave, I'm afraid I can't do that": Linguistics, Statistics, and Natural Language Processing circa 2001. To appear in the National Academies' Study on Fundamentals of Computer Science.

Lee, Lillian. "A non-programming introduction to computer science via NLP, IR, and AI." ACL workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, 2002.

Lee, Lillian. "Fast Context-Free Grammar Parsing Requires Fast Boolean Matrix Multiplication". *Journal of the ACM 49*, 2002.

Lee, Lillian; and Ando, Rie Kubota. "Iterative Residual Rescaling: An Analysis and Generalization of LSI". *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval (SIGIR)*, 2001.

Lee, Lillian; and Barzilay, Regina. "Bootstrapping Lexical Choice via Multiple-Sequence Alignment". *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.

Lee, Lillian; Pang, Bo; and Vaithyanathan, Shivakumar. "Thumbs up? Sentiment Classification using Machine Learning Techniques". *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.

Liben-Nowell, D., J. Kleinberg. The Link Prediction Problem for Social Networks. Proc. 12th International Conference on Information and Knowledge Management (CIKM), 2003.

MacKenzie, B. and S. B. Wicker, "Stability of Slotted Aloha with Multipacket Reception and Selfish Users," Proceedings of Infocom 2003. B. Krishnamachari, Y. Mourtada, and S. Wicker, "The Energy-Robustness Tradeoff for Routing in Wireless Sensor Networks," Proceedings of IEEE 2003 International Conference on Communications, ICC 2003, Anchorage, Alaska, May 2003.

Mateas, Michael; and Sengers, Phoebe, ed. "Narrative Intelligence". *Advances in Consciousness Series.* Amsterdam: John Benjamins Publishing Company, expected Fall 2002.

Meyerguz, L., D. Kempe, J. Kleinberg, R. Elber. The Evolutionary Capacity of Protein Structures. Proc. ACM RECOMB Intl. Conference on Computational Molecular Biology, 2004.

Mourtada, Y. and S. Wicker, "LGMA: Localized Gradient Management Algorithm for Mobile Wireless Sensor Networks," Symposium on Defense & Security, April 2004, Orlando, Florida.

Mourtada,Y., M. Swanson, and S. Wicker, "Statistical Performance Analysis of Address-Centric Performance versus Data-Centric Directed Diffusion Approach in Wireless Sensor Networks," Aerosense 2003 Symposium on Battlespace Digitization and Network Centric Systems III, April 2003.

Mourtada, Y. and S. Wicker, "On the Effect on Robustness of Topological Factors in Wireless Sensor Networks", Proceedings of the Workshop on Mobile and Wireless Networks MWN 2003, December 2002.

Ng, Vincent; and Cardie, Claire. "Combining Sample Selection and Error-Driven Pruning for Machine Learning of Coreference Rules". *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Ng, Vincent; and Cardie, Claire. "Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution". *Proceedings, COLING-2002.*

Ng, Vincent; and Cardie, Claire. “Improving Machine Learning Approaches to Coreference Resolution”. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics.

Ng, Vincent and Claire Cardie. Weakly Supervised Natural Language Learning Without Redundant Views. *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003),* 173–180, Association for Computational Linguistics, 2003.

Ng, Vincent and Claire Cardie. Bootstrapping Coreference Classifiers with Multiple Machine Learning Algorithms. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-2003),* Association for Computational Linguistics, 2003.

Ng, Vincent and Claire Cardie. Combining Sample Selection and Error-Driven Pruning for Machine Learning of Coreference Rules. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2002.

Penny, Simon; Smith, Jeffrey; Sengers, Phoebe; Bernhardt, Andre; and Schulte, Jamieson. "Traces: Embodied Immersive Interaction with Semi-Autonomous Avatars." *Convergence.* Vol. 7, No. 2, 2001.

Roberts, Nancy. "Using Bayesian Networks and Decision Theory to Model Physical Theory". A report for SUNY IT Masters degree, Dec. 2001.

Roth, M., S. Wicker, "Route Filtering in Swarm Intelligent MANETs," Tenth Annual International Conference on Mobile Computing and Networking (MobiCom), Philadelphia, USA, 2004.

Roth, M., S. Wicker, "Termite: Emergent Ad-Hoc Networking," The Second Mediterranean Workshop on Ad-Hoc Networks, Medhia, Tunisia, 2003.

Roth, M., S. Wicker, Termite: Ad-Hoc Networking with Stigmergy, IEEE Globecomm 2003, San Francisco, USA, 2003.

Sakk, E. and Wicker, S.B., Wavelet packets for error control coding, Proceedings of the SPIE 48th Annual Meeting (Wavelets X), San Diego, CA, August 3-8, 2003.

Samar, P. and S.B. Wicker, "On the Behavior of Communication Links in a Multi-Hop Mobile Environment," Frontiers in Distributed Sensor Networks, S.S. Iyengar and R.R. Brooks (eds.), CRC Press, 2004.

Samar, P. and S. B. Wicker, "Characterizing the Communication Links of a Node in a Mobile Ad Hoc Network," ICC 2004, Paris, France.

Samar, P. and S.B. Wicker, "On the Behavior of Communication Links of a Node in a Multi-Hop Mobile Environment," accepted for presentation at The Fifth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Tokyo, Japan, 2004.

Sellmann, Meinolf, The Practice of Approximated Consistency for Knapsack Constraints, Accepted at AAAI, 2004.

Sellmann, Meinolf and Torsten Fahle, Constraint Programming Based Lagrangian Relaxation for the Automatic Recording Problem, Annals of Operations Research (AOR). Vol. 118, pp. 17-33, 2003.

Sellmann, Meinolf, Approximated Consistency for Knapsack Constraints, Proceedings of the 9th intern. Conference on the Principles and Practice of Constraint Programming (CP), Springer LNCS 2833, pp. 679-693, 2003.

Sellmann, Meinolf, Cost-Based Filtering for Shorter Path Constraints, Proceedings of the 9th intern. Conference on the Principles and Practice of Constraint Programming (CP), Springer LNCS 2833, pp. 694-708, 2003.

Sellmann, Meinolf, Norbert Sensen, and Larissa Timajev, Multicommodity Flow Approximation used for exact Graph Partitioning, Proceedings of the 11th Annual European Symposium on Algorithms (ESA), Springer LNCS 2832, pp. 752-764, 2003.

Sellmann, Meinolf, Reduction Techniques in Constraint Programming and Combinatorial Optimization, Ph.D. Thesis, University of Paderborn, 2003.

Sengers, Phoebe, Joseph "Jofish" Kaye, Kirsten Boehner, Jeremiah Fairbank, Geri Gay, Yevgeniy Medynskiy, Susan Wyche. "Culturally Embedded Computing." Pervasive Computing, Vol 3, No 1, 2004.

Sengers, Phoebe, Kirsten Boehner, Geri Gay, Joseph "Jofish" Kaye, Michael Mateas, Bill Gaver, and Kristina Höök. "Experience as Interpretation." CHI 2004 Workshop on Cross-Dressing and Boundary Crossing: Exploring Experience Methods Across the Disciplines. Vienna, Austria, April 2004.

Sengers, Phoebe. "The Agents of McDonaldization." In Sabine Payr, ed., Agent Culture. Lawrence Erlbaum, in press, expected 2004.

Sengers, Phoebe. "Experiences in Designing Experiences." *CHI 2002 Workshop on Funology*, Minneapolis, MN, April 2002.

Sengers, Phoebe. "Narrative and Schizophrenia in Artificial Agents." *In Narrative Intelligence*, ed. Mateas and Sengers. Expected Fall 2002. An alternative version appeared in the *SigGraph 2001 Electronic Arts & Animation Catalog*. Another alternative version will appear in *Leonardo*, Vol 35, No 2, August 2002.

Sengers, Phoebe; Liesendahl, Rainer; Magar, Werner; Seibert, Christoph, Muller, Boris, Joachims, Thorsten, Geng, Weidong; Martensson, Pai; and Hook, Kristina. "The Enigmatics of Affect." *Conference on Designing Interactive Systems*, London, England, June 2002.

Stoyanov, Veselin, Claire Cardie, Janyce Wiebe, and Diane Litman, Evaluating an Opinion Annotation Scheme Using a New Multi-Perspective Question and Answer Corpus. *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, AAAI Press, to appear.

Swamy, C. and D.B. Shmoys. "Fault-tolerant facility location". Proceedings 14th Annual ACM-SIAM Symposium on Discrete Algorithms, 2003, 735-736.

Strogatz, S. The math of the real world. In: How a Child Becomes a Scientist (edited by John Brockman, Pantheon Books, New York, 2004).

Strogatz, S. 2003. Sync: The Emerging Science of Spontaneous Order. New York: Hyperion.

Strogatz, S. The real scientific hero of 1953. The New York Times, Op-Ed page, March 4 (2003).

Strogatz, S. How the blackout came to life. The New York Times, Op-Ed page, August 25 (2003).

Strogatz, S.H. Fermi's 'little discovery' and the future of chaos and complexity theory. In: The Next Fifty Years: Science in the First Half of the Twenty-First Century (edited by John Brockman, Vintage Books, New York, 2002).

Tsochantaridis, I., T. Hofmann, T. Joachims, and Y. Altun, Support Vector Machine Learning for Interdependent and Structured Output Spaces, Proceedings of the International Conference on Machine Learning (ICML), ACM Press, 2004.

Vetsikas, Ioannis A. and Bart Selman. A principled study of the design tradeoffs for autonomous trading agents. *Second International Joint Conference on Autonomous Agents and Multi-Agent Systems,* Melbourne, 2003.

Wei, Wei and Selman, Bart. Accelerating Random Walks. *Proceedings of 8th Intl. Conference on the Principles and Practice of Constraint Programming (CP-2002),* 2002.

White, Michael, Claire Cardie, Vincent Ng, and Daryl McCullough. Detecting Discrepancies in Numerical Estimates Using Multidocument Hypertext Summaries. *Proceedings of the Second International Conference on Human Language Technology Research (HLT-02)*, 2002.

White, Michael and Claire Cardie. Selecting Sentences for Multidocument Summaries Using Randomized Local Search. *ACL Workshop on Automatic Summarization*, 2002.

Wicker, S. B. and Kim, S., Fundamentals of Codes, Graphs, and Iterative Decoding, Boston: Kluwer Academic Press, 2002.

Wicker, S. B., "Cyclic Codes" in Wiley Encyclopedia of Telecommunications, (ed: J. Proakis), 2002.

Wiebe, Janyce, Eric Breck, Chris Buckley, Claire Cardie, Paul Davis, Bruce Fraser, Diane Litman, David Pierce, Ellen Riloff, Theresa Wilson, David Day, Mark Maybury. Recognizing and Organizing Opinions Expressed in the World Press. *2003 AAAI Spring Symposium on New Directions in Question Answering*, 12–19, AAAI Press, 2003.

Wiebe, Janyce, Eric Breck, Chris Buckley, Claire Cardie, Paul Davis, Bruce Fraser, Diane Litman, David Pierce, Ellen Riloff, Theresa Wilson. NRRC Summer Workshop on Multiple-Perspective Question Answering: Final Report. 2002.

Williams, R.; Gomes, C.; and Selman B., Backdoors To Typical Case Complexity, *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI03)*, 2003.

Williams, R.; Gomes, C.; and Selman B., On the connections between backdoors, restarts, and heavy-tails in combinatorial search, *Proceedings of the Sixth International Conference on Theory and Applications of Satisfiability Testing (SAT03)*, 2003.

Zhang, X. and S. B. Wicker "How to distribute sensors in a random field," IPSN 2004, Berkeley, California.

7. Honors/Awards

Claire Cardie

National Science Foundation Faculty Early CAREER Development Award, 1996-2000
Ralph S. Watts College of Engineering Excellence in Teaching Award, Cornell University, 1996
Lilly Teaching Fellow, Cornell University, 1996-1997
Best Written Paper Award, Ninth National Conference on Artificial Intelligence, Honorable Mention, 1991
Graduate Fellow, University of Massachusetts, 1993-1994
Massachusetts Regents Fellowship, University of Massachusetts, 1992-1993

Raffaello D'Andrea

RoboCup World Champions, 2002-2003

System architect and faculty advisor for the F180 League world champion Cornell Autonomous Robotic Soccer team. Fukuoka, Japan 2002.
Presidential Early Career Award for Scientists and Engineers (PECASE), nominated by AFOSR, 2001
Distinguished Lecturer, National Science Foundation Research Highlight Series, 2001
CAREER Award, National Science Foundation, 2000
J.P. and Mary Berger '50 Excellence in Teaching Award, 2000
Natural Sciences and Engineering Research Council of Canada 1967 Fellow, 1991-1996
University of Toronto Wilson Medal, 1991

Johannes Gehrke
Cornell University Provost's Award for Distinguished Scholarship (2004)
Alfred P. Sloan Foundation Fellowship (2003)
National Science Foundation CAREER Award, 2002
Cornell College of Engineering James and Mary Tien Excellence in Teaching Award, 2001
IBM Faculty Development Award 2000 and 2001

Carla Gomes
Elected to the Executive Council of the American Association for Artificial Intelligence (AAAI, over 6,000 members worldwide), the policy making body for AAAI. Direct vote by the AAAI members (2002-2005).

Selected Conference Chair for the International Conference on Principles and Practice of Constraint Programming (CP-2002), the premier conference in the field of constraint programming and constraint optimization methods.

Joe Halpern
Awarded Guggenheim and Fulbright Fellowships, 2001-02.
Selected Fellow of the Association for Computing Machinery, 2002.
Awarded 1997 Gödel Prize for outstanding paper in the area of theoretical computer science for ``Knowledge and common knowledge in a distributed environment".
Elected Fellow of the American Association of Artificial Intelligence, 1993.

Juris Hartmanis
Recipient of the Grand Medal of the Latvian Academy of Sciences (Lielo Medalu), 2001
CRA Distinguished Service Award, 2000
Honorary Doctor of Humane Letters Honoris Causa from the University of Missouri-Kansas City, 1998
Honorary Doctoral Degree, Dr.h.c., Univ of Dortmund, Germany, 1995
Bolzano Gold Medal of the Academy of Sciences, Czech Republic, 1995
ACM Fellow (charter member), 1994
ACM Turing Award, 1993
Humboldt Foundation Award for Senior US Scientists, 1993-1994
Fellow, American Academy of Arts and Sciences, 1992
Cornell Outstanding Educator Award, 1989-1991

Walter R. Read Professor of Engineering, 1981
Member, National Academy of Engineering, 1989
Clark Teaching Award, 1988-89
Fellow, American Assoc. for the Advancement of Science, 1981

John Hopcroft

Fellow of the Association for Computing Machinery, 1994
Member, National Academy of Engineering, 1989
Fellow of the Institute of Electrical and Electronics Engineering, 1987
Fellow of the American Association for the Advancement of Science, 1987
Fellow of the American Academy of Arts and Sciences, 1987
Association for Computing Machinery/A.M. Turing Award, 1986
National Science Foundation Graduate Fellow, 1961-1964

Jon Kleinberg

Best research paper award, ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2003National Academy of Sciences Award for Initiatives in Research, 2001
David and Lucile Packard Foundation Fellowship, 1999
Office of Naval Research Young Investigator Award, 1999
Alfred P. Sloan Research Fellowship, 1997
NSF Faculty Early Career Development Award, 1997
IBM Outstanding Innovation Award, 2002

Lillian Lee

Best paper award, Joint Meeting of the Human Language Technology conference and the North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL), 2004. (co-winner: Regina Barzilay)
Alfred P. Sloan Research Fellowship, 2002-2004
James and Mary Tien Excellence in Teaching Award, 2002
Stephen and Marilyn Miles Excellence in Teaching Award, 1999

Bart Selman

Fellow of the American Association for the Advancement of Science (AAAS), 2003
Fellow of American Association of Artificial Intelligence (AAAI), 2001
Cornell Outstanding Educator Award, 2001
Stephen '57 and Marilyn Miles Excellence in Teaching Award, 2002
Alfred P. Sloan Research Fellow, 1999-2000
NSF CAREER Award, 1998-2002
Elected to the Executive Council of the American Association for Artificial Intelligence, the policy making body for AAAI (1999-2002)

Phoebe Sengers

Cornell Faculty Innovation in Teaching Grant to redesign COM S/INFO 130, Web Design and Programming

National Science Foundation Career Award, 2002-2007
Fulbright Fellowship, 1998-1999
Office of Naval Research Allen Newell Graduate Fellowship, 1994-1997
National Science Foundation Graduate Research Fellowship, 1990-1993

Christine Shoemaker
Humboldt Research Prize for Senior Scientists, 2002-2003
Elected a Fellow of ASCE, 1996

Steve Strogatz
MIT E. M. Baker Award for Excellence in Undergraduate Teaching
NSF Presidential Young Investigator Award